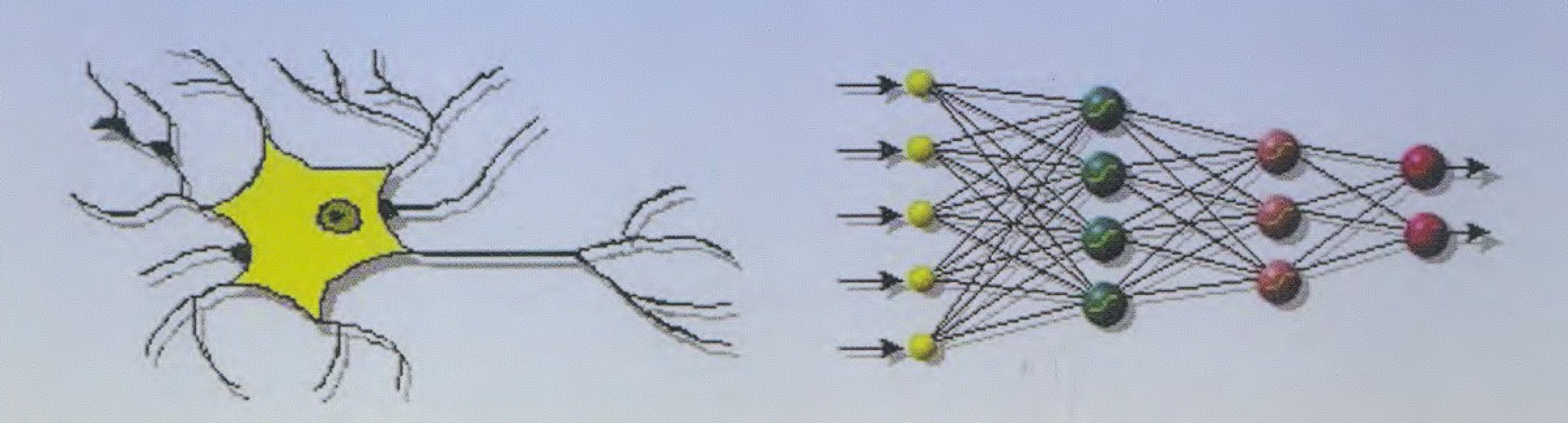



# الشبكات العصبونية الصنعية
## بين النظرية والتطبيق

***الجزء الثاني***



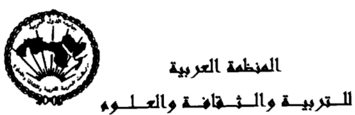
إعداد

الدكتور المهندس أحمد الكرمو

مراجعة

الدكتور المهندس راكان رزوق

الدكتور المهندس رامز حاج إسلام



المنظمة العربية
للتربية والثقافة والعلوم

المركز العربي
للتعريب والترجمة والتأليف والنشر

# الشبكات العصبونية الصنعية

## بين النظرية والتطبيق

إهـــــــداء٢٠٠٨
المركز العربى للتعريب والترجمة والتاليف والنشر
الجمهورية العربية السورية

# الشبكات العصبونية الصنعية

## بين النظرية والتطبيق

### *الجزء الثاني*

إعداد

**الدكتور المهندس أحمد الكرمو**

مراجعة

**د.م. راكـــان رزوق** **د.م. رامز حاج إسلام**

دمشق 2001

**الشبكات العصبونية الصنعية بين النظرية والتطبيق - الجزء الثاني**

تأليف: د. أحمد الكرمو

المركز العربي للتعريب والترجمة والتأليف والنشر بدمشق

ص.ب: 3752 ـــ دمشق ـــ الجمهورية العربية السورية

هاتف: 3334876 11 963 + ــ فاكس: 3330998

E-mail: acatap@net.sy

Web Site: www.acatap.htmlplanet.com

# المحتويات

**مقدمة**

**الفصل الثامن : الشبكات التكرارية الديناميكية**

**الفصل التاسع: آلات بولتزمان ومحاكاة التلدين**



**الفصل العاشر: الشبكات العصبونية الصنعية ذاتية النمو**

**الفصل الحادي عشر: شبكات النيوكونتيرون**



**الفصل الثانـــي عشر: الشبكات المبنية احتمالياً**

## الفصل الثالث عشر: خريطة الملامح الذاتية التنظيم والتكميم الشعاعي

## الفصل الرابع عشر :نظرية الطنين المتكيف



## الفصل الخامس عشر: الأنظمة العصبونية العائمة، الحساب المرن، الخوارزميات الوراثية، شبكات المنطق العصبوني

## الفصل الثامن

# الشبكات التكرارية الديناميكية
# Dynamic Recurrent Networks

في هذا الفصل سنتعرض لصنف خاص من الشبكات الديناميكية يشمل الشبكات العصبونية الصنعية التكرارية RNNs (Recurrent Neural Networks). تختلف الشبكات العصبونية التكرارية عن أنواع شبكات هوبفيلد الموصوفة في الفصل الخامس في إمكانية أن يكون لها طبقات متعددة، ومصفوفة أوزان ليست متناظرة، وأيضاً تغذية عكسية ذاتية، ومن الممكن استعمال خوارزميات التعليم بمعلم للانتشار الخلفي لتدريب هذه الشبكات.

سنبدأ بوصف عام لهذه الشبكات وديناميكيتها، بعد ذلك سنركز على شبكات الانتشار الخلفي التكرارية وبعض حالاتها الخاصة، وسنتوقف عند بعض بناها الخاصة وقدراتها الحسابية. أخيراً سنلخص الفصل ببعض التطبيقات النموذجية لهذه الشبكات التكرارية الديناميكية.

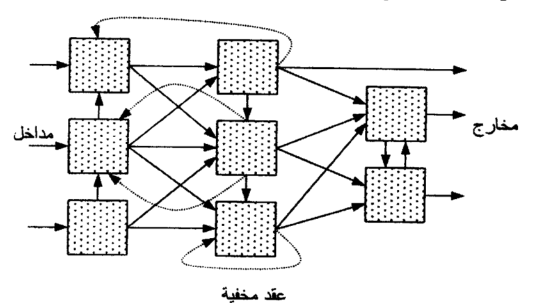
### 1.8 تمهيد

يمكن تعميم شبكات التغذية الأمامية الموصوفة في الفصول السابقة لتعمل بأسلوب تكراري، وذلك بوصل خرج عنصر معالجة واحد أو أكثر إلى مدخل أو أكثر لعناصر المعالجة الموجودة في نفس الطبقة أو الطبقات السابقة. أيضاً، يمكن أن يكون لهذه البنية المعممة وصلات جانبية بين الوحدات في نفس الطبقة بما في ذلك وصلات التغذية العكسية الذاتية. إن دمج وصلات التغذية العكسية مع الشبكات المتعددة الطبقات الأمامية التغذية يحدث تغيراً كبيراً في عمل وخوارزميات تعليم هذه الشبكات مقارنة مع مقابلاتها من الشبكات غير الديناميكية (الساكنة). يمكن أن يكون لبنى الشبكات التكرارية الديناميكية وصلات

جانبية ضمن طبقة معطاة بالإضافة إلى الوصلات التـــي تغذي عكسياً الإشارات إلى الطبقات المتتالية بما في ذلك طبقة الخرج.

لن نطيل التفكير في أن الشبكة لها طبقات عديدة، وبدلاً من ذلك سننظر ببساطة إلى هذه الشبكات على أنها مؤلفة من عدد من وحدات المعالجة الموصلة داخلياً، حيث يمكن أن توصل أية وحدة i إلى أية وحدة j بما في ذلك حالة i = j. ويمكن أن تعتبر أية وحدة كوحدة خرج، على حين ستستقبل بعض الوحدات الخاصة المداخل بما في ذلك مداخل الانحياز.

إذا كان للشبكة n وحدة تعمل منها m وحدة لاستقبال الدخل الخارجي فإنه يمكن أن تستعمل مصفوفة الأوزان الوحيدة **W** ببعد $n \times (n + m)$ لتعيين وسطاء الأوزان الليفية على نحو كامل للشبكة. تستقبل الوحدات ذات الوصلات الحلقية إشارة تغذية عكسية من نفس




**الشكل 1.8**: شبكة عصبونية تكرارية عامة

الوحدة؛ وهذا يعنـــي، أن الإشارات ستكون على الأقل معتمدة جزئياً على الحسابات السابقة المنفذة بنفس الوحدة. يوضح الشكل (1.8) بنية شبكة عصبونية تكرارية عامة.

يبدو من هذا الشكل أن للشبكة ثلاثة مداخل خارجية وثلاثة مخارج وعدداً من الوصلات الداخلية بين الوحدات بما في ذلك خطوط التغذية العكسية الذاتية وغير الذاتية. رأينا فيما سبق أن الشبكات العصبونية كشبكة هوبفيلد لها أوجه شبه فيزيائية يمكن أن تكون أحياناً

مفيدة في وصف هذه الشبكات وفهم سلوكها. يمكن من وجهة النظر هذه أن يفهم أكثر سلوك الشبكات التكرارية الديناميكية بمعرفة الظاهرة الديناميكية غير الخطية، مثل تدفق السائل الدوراني أو أنظمة التحكم غير الخطية.

يمكن أن توصف هذه السلوكية وصفاً كاملاً بواسطة مجموعة من المعادلات التفاضلية غير الخطية من المرتبة الأولى من الشكل:

$$\frac{dx_i}{dt} = G_i(\mathbf{W}, \mathbf{I}, \mathbf{x}(t)), \quad i = 1,2,\ldots,n \qquad (1.8)$$

حيث **x** شعاع الحالة، و**W** مصفوفة قوى الوصل الليفية (الأوزان)، و**I** شعاع دخل خارجي، و$G_i$ تابع تفاضلي غير خطي. في نموذج الزمن المتقطع، تكون ديناميكيات الشبكة موصوفة بمجموعة من معادلات الفروق غير الخطية من الشكل:

$$x_i\,(t+1) = G_i(\mathbf{W}, \mathbf{I}, \mathbf{x}(t)) \qquad (2.8)$$

حيث $x_i(t)$ خرج الوحدة رقم $i$ عند اللحظة $t$. في أنظمة التغذية العكسية يصبح السؤال عـــن الاستقرار أمراً هاماً، ويحدد سلوك الشبكة التكرارية الديناميكية بواسطة الوسطاء **W**، و**I**، والشروط الأولية أو نقطة البداية $\mathbf{x}(0)$. يمكن أن يتطور هذا السلوك بإحدى الطرائق التالية:

1. التقارب إلى نقطة جذب مستقرة.
2. الاستقرار إلى اهتزازات حلقية دورية.
3. الميل باتجاه سلوك شبه دوري (اهتزازات عند ترددات معينة).
4. تبدي الشبكة نوعاً من السلوك التائه الفوضوي (الموصوف في الفصل الثالث).

في البداية سنركز على حالة التقارب المستقر حيث تتقارب الشبكة إلى نقطة جذب وحيدة أو تنجز تطبيقاً ما مرغوباً به على شعاع الدخل، وسنذكر بعض الشروط الكامنة خلف عدم التأكد من هذا السلوك المستقر.

لكي نرى كيف تعمل الشبكة التكرارية الديناميكية سنفترض أن الشبكة الموضحة في الشكل (1.8) عملت لبعض الوقت في الزمن المتقطع. عند اللحظة $t > 0$ يقدم نموذج الدخل **x** إلى وحدات دخل الشبكة، ومن ثم تقوم هذه الوحدات بحساب تفعيلها $\mathbf{f_i}(\mathbf{x}, \mathbf{W})$ وتمريره إلى جميع الوحدات المتصلة معها في اللحظة الزمنية $t + 1$. تقوم الوحدات الأخرى أيضاً (غير

وحدات المداخل) بحساب تفعيلها وتمريره إلى جميع الوحدات المتصلة معها في اللحظة الزمنية t + 1. تتكرر هذه العملية عند كل لحظة زمنية t + i حيث $i = 1,2,..,n$.

يجري في كل خطوة زمنية حساب الدخل التركيبـــي net لكل وحدة من إشارات دخل التغذية الأمامية والتغذية العكسية لها. قد تتقارب قيم تفعيل وحدات الخرج لتستقر عند نقطة ثابتة، أو تهتز لتكرر نفس مجموعة القيم المختلفة عبر الزمن، أو تعطي نوعاً من الشرود الفوضوي وذلك بالاعتماد على قيم شعاع الدخل **x** ووسطاء الشبكة الأخرى.

الديناميكيات الفعلية للشبكات التكرارية الديناميكية المستمرة (ستوصف فيما بعد) مشابهة لديناميكيات شبكات هوبفيلد المستمرة الموصوفة في الفصل الخامس. يمكن إثبات أن الشبكات التكرارية الديناميكية هي تعميم لشبكات التغذية الأمامية وذلك باستنتاج شبكة متعددة الطبقات أمامية التغذية مكافئة لشبكة التكرار الديناميكية (Minsky & papert عام 1969 [90]).

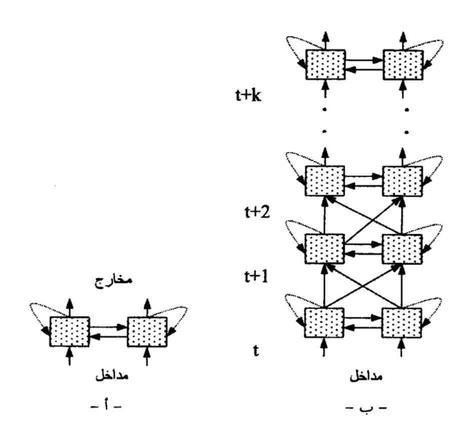
نقول عن شبكتين إنهما متكافئتان عندما تبديان نفس السلوك. وهذا يمكن أن ينجز من خلال عملية النشر في الزمن، حيث توافق كل خطوة زمنية لشبكة التكرار الديناميكية طبقةً إضافيةً في الشبكة المتعددة الطبقات الأمامية التغذية.

يوضح الشكل (2.8) شبكة MLFF مكافئة لشبكة تكرار ديناميكية بسيطة بوحدتين متصلتين اتصالاً كاملاً (Rumelhart عام 1986 [56]).

الشبكة MLFF المكافئة (الشكل (2.8 –ب–)) لشبكة التكرار الديناميكية بوحدتين متصلتين اتصالاً كاملاً (الشكل (2.8–أ–)) لها نفس قيم الأوزان في جميع الطبقات (وهي نفس قيم الأوزان على وصلات وحدتـــي شبكة التكرار الديناميكية)، ولكن مخارج العقد تختلف على كل الطبقات المتتابعة.

تستطيع مثل هذه الشبكات حل مسألة XOR غير الخطية، هذه المسألة المستحيلة الحل بالنسبة لشبكة MLFF بعقدتين فقط. من الواضح أنه يمكن استعمال خوارزمية الانتشار الخلفي للخطأ لتدريب هذه الشبكات.

لقد استعملت هذه الشبكات في عدد من التطبيقات الهامة مثل الذواكر المترافقة، وتعرّف إشارة الكلام، والتحكم، والتفضيل، والتنبؤ، وتوليد متتاليات النماذج.




**الشكل 2.8:** ـــ أ ـــ شبكة تكرار بسيطة متصلة بالكامل ـــ ب ـــ شبكة MLFF منشورة في الزمن مكافئة

## 2.8 ديناميكيات الشبكات العصبونية التكرارية العامة

## The Dynamics of General Recurrent Networks

كما في العديد من الأنظمة الديناميكية غير الخطية الأخرى، ما يزال سلوك الشبكات العصبونية التكرارية العامة موضوعاً هاماً مطروحاً للبحث، ومع ذلك جرت عدة دراسات باستعمال بنـــى مختلفة وخوارزميات تعليم متطورة ناجحة.

استُعملت الشبكات التكرارية عموماً لتقوم بنوعين أساسيين من الحسابات: النوع الأول مـــن المهام هـــو تعلـــم إنجاز تطبيـــق عـــام $f: R^n \rightarrow R^m$ متعلق مع الزمن أيضاً، أي: $y(t) = f[x(t)]$. ينجز التطبيق $x \rightarrow y$ عملياً بعدد من اللحظات الزمنية المتقطعة ضمن المجال $[t_0, t_k]$. وهكذا، إذا كانت الحالة الأولية لخرج النظام هي $y(t_0)$، فإن هذا النظام سيطبق مجموعـــة النماذج المؤقتة التالية: $\{x(t_0), x(t_1), .., x(t_{k-1})\}$ فـــي فراغ الدخـــل إلى

المجموعة: $\{y(t_1), y(t_2), y(t_3), ..., x(t_k)\}$ في فراغ الخرج. وستكون هذه التطبيقات واحداً لواحد بالتقابل على فراغ الخرج، مفيدةً في مسائل التحكم والتصنيف والتنبؤ حيث يكون التعميم ضرورياً.

يمكن استعمال العديد من هذه التطبيقات لكل نموذج (كثير لواحد) عندما لا تعطي تغيرات صغيرة في قيم الدخل تغيراً في الخرج. وهذا النمط من الحساب مفيدٌ في مسائل الاستدعاء المترافق وارتباط الخطأ. واستُعملت أيضاً لنمذجة سلوك الأنظمة الفوضوية والاهتزازية، ولكننا لن نتعرض إلى هذه الديناميكيات في هذا الفصل.

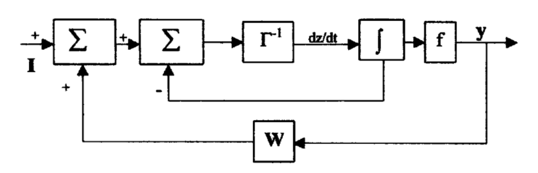
لقد دَرَس سلوك الشبكات العصبونية الصنعية التكرارية كثير من الباحثين مثل Pineda عام 1987 [91] و1988 [92] و1989 [93] وAlmeda عام 1987 [94] وRumelhart وزملاؤه عام 1986 [56] وWilliams & Zipser عام 1989 [95] وpearlmutter عام 1988 [96] وZipser عام 1989 [97] وWilliams & Peng عام 1990 [98] وغيرهم.

لقد رأينا من قبل في الفصل الخامس نموذجاً عن الشبكات التكرارية جرى عرضه من خلال شبكات هوبفيلد. ونذكر هنا أن شبكة هوبفيلد هي نوع خاص من الشبكات التكرارية بمصفوفة وزن متناظرة **W** وبدون خطوط تغذية عكسية ذاتية، أي . توصف ديناميكية شبكات هوبفيلد المستمرة بواسطة معادلات تفاضلية غير خطية مترابطة (المعادلات (31.5) وسنكررها هنا بالشكل المناسب) كما يلي:

$$\tau_i \frac{dz_i(t)}{dt} = -z_i(t) + \sum_{j=1}^{n} w_{ij} y_j + I_i \quad ,i = 1,2,..,n \qquad (3.8)$$
$$y_i(t) = f[z_i(t)]$$

حيث $\tau_i$ أزمنة الاسترخاء (relaxation)، و$z_i$ الحالات الداخلية للوحدة i، و $y_i$ حالة الخرج (قيمة التفعيل) للوحدة i، و $w_{ij}$ وزن الوصلة من الوحدة j إلى الوحدة i، و$I_i$ هي المداخل الخارجية بما في ذلك الانحيازات للوحدة i. يوضح الشكل (3.8) المخطط الصندوقي لتدفق هذا النظام. يمثل الصندوق $\Gamma^{-1}$ في هذا المخطط الثابت الزمني اللازم لتحقيق المعادلة (1.8). المخططات الصندوقية كهذه توضح بجلاء تأثير وصلات التغذية العكسية وتري كيف

ينـــى ببساطة نموذج المحاكاة المستمر للشبكة. نُذكّر بأن مصفوفة الأوزان المتناظرة في حالة هوبفيلد شرطٌ كافٍ لضمان سلوك مقارب مستقر (أي لكي يكون التقارب إلى نقطة ثابتة مستقرة). لقد كررنا إعطاء المعادلات والمخطط الصندوقي هنا لشبكة هوبفيلد لأن ديناميكيات هذه الشبكة يمكن مقارنتها مع ديناميكيات الشبكة العصبونية التكرارية التـــي سنناقشها فيما يلي.




**الشكل 3.8:** مخطط تدفق شبكة هوبفيلد المستمرة

## 1.2.8 ديناميكيات الشبكات العصبونية التكرارية المستمرة

**Continuous Recurrent Neural Network Dynamics**

إن ديناميكية الشبكة العصبونية التكرارية في الزمن المستمر مشابهة لديناميكية شبكات هوبفيلد، وتوصف عموماً بمعادلات تفاضلية غير خطية مترابطة من الشكل:

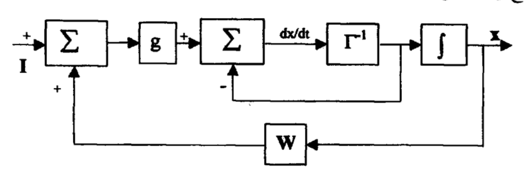
$$\tau_i \frac{dx_i(t)}{dt} = -x_i(t) + g_i\left(\sum_{j=1}^{n} w_{ij}x_j + I_i\right) \quad ,i = 1,2,\ldots,n \qquad (4.8)$$

حيث $\tau_i$ عوامل الاسترخاء الزمنية، و $x_i(t)$ حالة الوحدة رقم i عند اللحظة t، و $w_{ij}$ وزن الوصلة من الوحدة j إلى الوحدة i، و $I_i$ هو الدخل الخارجي للوحدة i، و $G_i$ توابع تفعيل غير خطية. لاحظ أن المعادلة (4.8) مرتبطة مع المعادلة (3.8) من خلال تحويل خطي بسيط:

$$\mathbf{z(t) = w\, x(t) + I} \qquad (5.8)$$

حيث **z(t)** هو شعاع حالة شبكة هوبفيلد، و**W** مصفوفة الوزن (القابلة للقلب)، و**x(t)** شعاع الحالة للشبكة التكرارية الديناميكية العامة، و**I** شعاع الدخل. يبين مخطط التدفق

الموضح في الشكل (4.8) الفرق بين هذه الشبكات وشبكات هوبفيلد.




**الشكل 4.8**: مخطط تدفق شبكة تكرارية عامة مستمرة

يستطيع المرء، بوجه خاص، مقارنة مكان توابع التفعيل f و g وخطيّ التغذية العكسية في المخططين. سنحدد الحالة الأولية للنظام بواسطة $\mathbf{x}^{0}$ والحالة النهائية الثابتة على وجه التقريب بواسطة $\mathbf{x}^{f}$. نقاط الجذب المستقلة زمنياً هي حلول معادلات الحالة الثابتة (أي dx/dt = 0):

$$x_i^f = g_i\left(\sum_{j=1}^{n} w_{ij} x_i^f + I_i\right) \quad ,i = 1,2,\dots,n \qquad \textbf{(6.8)}$$

تميز الأنظمة ذات الديناميكية الموصوفة بالمعادلة (3.8) بثلاث خواص هامة:

1. أنها تعمل بعدة درجات من الحرية؛ توافق درجات الحرية عدد مستويات التفعيل والمشتقات الزمنية للمستويات التـــي تكوّن فراغ طور النظام.
2. ديناميكيتها غير خطية، نتيجة لتوابع التفعيل غير الخطية $g_i$ التـــي تُختار أساساً لمقدرات التطبيق العامة المنجزة بواسطة هذه الشبكات.
3. أنها تبديدية، أي يتقارب حجم فراغ الطور على جملة مولدة صغيرة الأبعاد.

تعطي هذه المميزات الثلاث لبعض الأنظمة سلوكاً غير عادي بالإضافة إلى مقدرتها الحسابية كما ذكر من قبل.

بنظرة عامة على النماذج التكرارية العامة، نستطيع التمييز بين ثلاثة أصناف من ديناميكيات الشبكات المعتمدة على قيم مصفوفة الوزن **W**:

1. في حال كون مصفوفة الأوزان مثلثية سفلى، ستكون الشبكة MLFF متعددة الطبقات أمامية التغذية عادية بدون خطوط تغذية عكسية كتلك التـــي درست في الفصول

السابقة.

2. أما في حال كون مصفوفة الأوزان **W** متناظرة بقطر رئيسي صفري، فإن الشبكة ستكون تكرارية ديناميكية من نوع هوبفيلد، وستكون مضمونة التقارب إلى واحد من عدة جواذب نقطة ثابتة في حال دراسة تابع (طاقة) **Lyapunov**.
3. أما عندما تكون مصفوفة الأوزان **W** غير متناظرة عامة، فإن الشبكة ستكون تكرارية ديناميكية والتقارب لن يكون مؤكداً ما لم تفرض بعض الشروط المقيدة على مصفوفة الأوزان.

في الحالة العامة لشبكة التكرار الديناميكي، سيتطور النظام إلى إحدى الطرائق المذكورة آنفاً بالاعتماد على وسطاء النظام وعلى الحالة الأولية، أي يمكن أن يتقارب لواحد من حالات الاستقرار المحدودة (جواذب النقطة الثابتة على نحوٍ مشابه لشبكات هوبفيلد)، أو أنه يبدي نوعاً من السلوك المهتز، أو يذهب إلى السلوك التائه الفوضوي (اهتزاز عند عدد غير محدود من الترددات). وكما ذكرنا فإن السلوك الذي تبديه شبكة خاصة سيعتمد على الوصلات الليفية وقيم وسطاء الشبكة والحالة الأولية، وقد ثبت أن سلوك الاستقرار سيكون مضموناً إذا كان مطال الأوزان محدداً (Atia عام 1988 [99]).

مثلاً، إذا كان g′ هو مشتق g تابع التفعيل التفاضلي غير الخطي $(g' = dg(x)/dx)$ و $w_{ij}$ الأوزان الواصلة للوحدة j مع الوحدة i فإن تقارب النظام سيكون مؤكداً إلى نقطة جذب وحيدة إذا تحققت هذه المتراجحة:

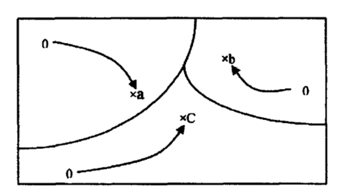
$$\sum_i \sum_j w_{ij}^2 < \frac{1}{\left(\max_i |g_i'|\right)^2} \qquad (7.8)$$

في هذه الفقرة سينصب اهتمامنا على حالة التقارب.

في الحالة المستقرة سيكون هناك طريقتان للعمل بالاعتماد على طريقة دخل المعطيات إلى الشبكة: التطبيقات المستمرة أو التقارب لواحدة من عدة نقاط جذب ثابتة. تحدِّد كلا الطريقتين بواسطة نوع معطيات الدخل إلى الشبكة؛ إما بواسطة وضع قيم الحالة الأولية $\mathbf{x}^0$ وإما بواسطة تخصيص مداخل خارجية **I** (الانحيازات ونماذج الدخل). سنشير لكلتا حالتي العمل بمطبق النقطة الثابتة والمطبق المستمر على الترتيب (وقد سمّاهما **Pineda** دخل الحالة

الأولية والدخل الوسيطي على الترتيب عام 1988 [92]).

عندما تمثل الحالة الأولية $\mathbf{x}^0$ دخل الشبكة و$\mathbf{I}$ معطى كقيمة ثابتة ما لكل أزمنة الاسترخاء، فإن شبكة التكرار الديناميكي ستحسب تطبيق M من الحالة الأولية $\mathbf{x}^0$ إلى الحالة النهائية $\mathbf{x}^t$؛ أي $\mathbf{M}: \mathbf{x}^0 \rightarrow \mathbf{x}^t$. إن النقاط الأولية $\mathbf{x}^0$ الواقعة ضمن حوض التجاذب ستتقارب إلى نقطة ثابتة ضمن هذا الحوض كما هو موضح في الشكل (5.8).




**الشكل 5.8**: النقاط الثابتة a, b, c وحدود حوضها

هذا هو نوع التطبيق (كثير لواحد) المنجز بواسطة الشبكة، المفيد في مسائل الاستدعاء المترافق وتصحيح الخطأ حيث يكون نموذج الدخل ناقصاً أو مشوباً بالضجيج.

إن سلوك المطبق المستمر يكون محققاً عندما تعامل المداخل الخارجية $\mathbf{I}$ كدخل للشبكة وتكون الحالة الأولية $\mathbf{X}^0$ لكل وحدات الدخل هي مجموعة من القيم الثابتة. في هذه الحالة، الحساب المنجز هو تطبيق من الدخل الوسيطي $\mathbf{I}$ إلى الحالة النهائية $\mathbf{x}^t$؛ أي $\mathbf{M}: \mathbf{I} \rightarrow \mathbf{x}^t$. هذا هو نوع التطبيق واحد لواحد ويكون ناعماً لأن تغيراً صغيراً في $\mathbf{I}$ يعطي تغيراً صغيراً في $\mathbf{x}^t$.

هذا التطبيق مشابه للتطبيقات المنجزة بواسطة الشبكات المتعددة الطبقات الأمامية التغذية، وهذه التطبيقات مفيدة لإنجاز التطبيقات العامة حيث يكون التعميم ضرورياً مثل قضية التحكم بالربوت ومسائل التصنيف والتنبؤ بمعطيات السلسلة الزمنية.

### 2.2.8 ديناميكيات الشبكة المتقطعة Discrete Network Dynamics

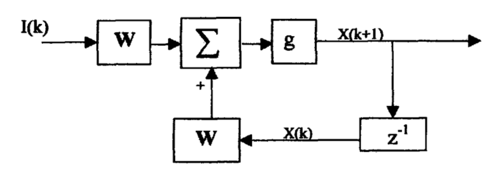
يمكن أن نحصل على ديناميكيات الزمن المتقطع لشبكات التكرار الديناميكية مـــن المعادلة (3.8) بوضع $\tau_i = 1$ والقيام بالتعويضات:

$$\frac{dx}{dt}=\frac{\Delta x}{\Delta t}=\frac{x(t+\Delta t)-x(t)}{(t+\Delta t)-t}$$

وليكن $\Delta t \to 1$ فنحصل على :

$$x_i(t+1)=g\left(\sum_{j=1}^{n} w_{ij}x_j(t)+I_j\right) \qquad (8.8)$$

يوضح الشكل (6.8) مخطط تدفق النظام الموافق للحالة المتقطعة، حيث يمثل الصندوق المتضمن $z^{-1}$ وحدة تأخير لإشارة التغذية العكسية.




**الشكل 6.8**: نموذج مخطط تدفق شبكة تكرارية متقطعة

لاحظ الفرق بين نماذج شبكة التكرار المستمرة والمتقطعة، ففي النموذج المستمر هناك ممران للتغذية العكسية مقابل ممر واحد في الحالة المتقطعة، ونحتاج إلى مصفوفتـــي أوزان في النموذج المتقطع لوصف سلوك الشبكة بدقة.

سيكون السلوك العملي للشبكتين نفسه إذا اخترت الزيادة الزمنية صغيرة ($\Delta t \to 0$) في حالة الزمن المتقطع. من ناحية أخرى، سيكون سلوك كلتا الشبكتين مختلفاً تماماً إذا اختير مقياس زمن المحاكاة كبيراً . فمثلاً، في حالة شبكتين إحداهما متقطعة والأخرى مستمرة بنفس الوسطاء، نجد أن الأولى قد تتقارب على حين أن الأخرى قمتز إذا لم يكن اختيار الفرق الزمنـــي بين الشبكتين صغيراً كفاية.

## 3.8 تدريب الشبكات العصبونية التكرارية

## Training Recurrent Networks

اقتُرح عدد من خوارزميات التعليم للبنـــى المختلفة لشبكات التكرار الديناميكية، وقد

أمكن تدريب الكثير من هذه البنـــى بخوارزمية الانتشار الخلفي مع بعض التعديلات البسيطة. فمثلاً، في بنـــى الشبكات التكرارية الخاصة التـــي لها خطوط تغذية عكسية جزئية فقط، ستعمل خوارزمية الانتشار الخلفي التقليدي بدون تعديلات. وأيضاً في حالة الشبكة الموضحة في الشكل (2.8)، يمكن أن يُستعمل نموذج بسيط "لخوارزمية الانتشار الخلفي خلال الزمن".

تتطلب هذه الطريقة أن تكون أخطاء وصلات التغذية العكسية الذاتية أيضاً مكدسة ومشتملة في عملية تعديل الأوزان. هناك تعديلات أخرى على خوارزمية الانتشار الخلفي ليست بسيطة كما سبق، فهي تتطلب جمع وتخزين الأخطاء عبر كل المسارات قبل أن ينفذ تحديث الأوزان بتدرج الهبوط.

سنناقش في هذا المقطع بعض أهم الطرق المقترحة، وسنشتق بعض التعميمات من لأشكال الانتشار الخلفي لنماذج الزمن المتقطع والمستمر في خوارزميات تعليم الشبكات العصبونية التكرارية. يمكن أن تعمم خوارزمية الانتشار الخلفي لتدريب الشبكات التكرارية العامـــة لتعمل كمطبق مستمر (continuous mapper) أو كمطبـــق نقطـــة ثابتـــة (fixed-point mapper) (Pineda عام 1987 [91]، وعام 1988 [92]، وعام 1989 [93]). الآن سنشتق خوارزمية للمطبق المستمر وسنحيل القارئ إلى مكان وجـــود الأخرى في المراجع.

نستطيع استعمال المعادلة (6.8) عندما يكون للشبكة مخارج مستقرة (نقاط ثابتة) لاشتقاق قاعدة خطأ الانتشار الخلفي المناسبة المبنية على تدرج الهبوط. لذلك سنتجاهل تأثير أي تأخيرات زمنية للانتشار في ديناميكيات الشبكة.

ما نحتاج إليه الآن هو إيجاد حل لمعادلة تحديث الأوزان التالية:

$$\Delta w_{ij} = -\eta \frac{\partial E}{\partial w_{ij}} = \eta \sum_k E_k \frac{\partial x_k^f}{\partial w_{ij}} \qquad (9.8)$$

حيث إن E الخطأ الكلي هو مجموع الأخطاء المربعة للعقد منفردة:

$$E = \frac{1}{2} \sum_{k=1}^{n} E_k^2 \qquad (10.8)$$

إذا كانت $x_k^f$ وحدة خرج فإن $E_k = d_k - x_k^f$ وإلا فإن $E_k = 0$ .

بإسقاط الدليل f لسهولة التعابير ووضع توابع التفعيل $g_i = g$ نجد باشتقاق معادلة الحالة المستمرة (المعادلة (6.8)):

$$\left(H_i \frac{\partial x_k}{\partial w_{ij}} = g'\right)\sum_1 \left\{\frac{\partial x_{k1}}{\partial w_{ij}} x_1 + w_{k1} \frac{\partial x_1}{\partial w_{ij}}\right\}$$

$$(H_i = g')\left[\delta_{kj} x_j + \sum_1 w_{k1} \frac{\partial x_1}{\partial w_{ij}}\right] \qquad (11.8)$$

حيث
$$\mathbf{H}_i = \sum_{j=1}^{n} w_{ij} x_j + I_i$$

نلاحظ من المعادلة (11.8) أن الحد الأول في القوس $\partial_{kj} x_j$ أتـــى من تبسيط المجاميع لأن العناصر في $\mathbf{W}$ مستقلة بالفرض، و $\delta_{kj}$ الذي يمثل عناصر القطر الرئيسي للمصفوفة الواحدية يمكن أن يستبدل بحدود المشتقات الجزئية لأنها ذات قيمة مساوية للواحد إذا وفقط إذا كان $k = i$ و $l = j$، وللصفر ماعدا ذلك.

نستطيع كتابة الطرف الأيسر للمعادلة (11.8) كمجموع جداءات مصفوفة واحدية مع المشتق الجزئي لـــ $x_l$ كما يلي:

$$\frac{\partial x_k}{\partial w_{ij}} = \sum_l \delta_{k1} \frac{\partial x_l}{\partial w_{ij}}$$

وبالتعويض والجمع لكل حدود الاشتقاق على الطرف الأيمن للمعادلة (10.8) يمكن أن نعيد كتابتها كما يلي:

$$\delta_{kj} x_j g'(H_i) = \sum_l L_{kl} \frac{\partial x_l}{\partial w_{ij}} \qquad (12.8)$$

حيث
$$L_{kl} = \delta_{kl} - g'(H_k) w_{kl}$$

لتكن $\mathbf{L}$ مصفوفة بعناصر $L_{kl}$ وليكن $\mathbf{L}^{-1}$ هو مقلوب المصفوفة $\mathbf{L}$. بضرب كلا طرفي المعادلة (12.8) بـــ $(\mathbf{L}^{-1})_{lk}$ نحصل على:

$$(\mathbf{L}^{-1})_{lk}\, xjg'(H_k) = \frac{\partial x_1}{\partial w_{ij}} \qquad (13.8)$$

وهذه المعادلة تعتبر حلاً أول يستعمل لتحديث الأوزان (9.8).

ولكن هذا الحل لسوء الحظ يتطلب قلباً للمصفوفة، ولتجنب هذه الحسابات الإضافية يمكن أن نعطي نظاماً ديناميكياً مرافقاً بوضع:

$$y_i = g'(H_i)\sum_r E_r(\mathbf{L}^{-1})_{ri}$$

وبحل هذه المعادلات فيما يتعلق بـــ $\mathbf{E_r}$ نحصل على:

$$E_r = \sum_i L_{ir}\left\{\frac{y_i}{g'(H_i)}\right\}$$

وبضرب كلا طرفي المعادلة بـــ $g'(H_r)$ وتعويض الشكل الصريح لـــ $\mathbf{L}$ والجمع عبر $\mathbf{r}$، عندها سنحصل على الشكل الشائع للتعبير:

$$0 = -y_r + g'(H_r)\left\{\sum_i w_{ir} y_i + E_r\right\} \qquad (14.8)$$

سيدرك المرء فعلاً أن حلول هذه المعادلة الخطية هي نقاط ثابتة لمعادلة تفاضلية موافقة من الشكل:

$$\frac{dy_r}{dt} = -y_r + g'(H_r)\left\{\sum_i w_{ir} y_i + E_r\right\} \qquad (15.8)$$

وبهدف تلخيص ما سبق سنعطي المعادلات التـــي تصف الديناميكية الكاملة لشبكة المطبق مستمر:

$$\tau_i \frac{dx_i}{dt} = -x_i + g_i\left(\sum_{j=1}^{n} w_{ij} x_j + I_i\right) \quad ,i = 1,2,\ldots,n \qquad (16.8)$$

$$\tau_i \frac{dw_{ij}}{dt} = x_j g'(H_i)\sum_r E_r(\mathbf{L}-1)_{ri} \qquad (17.8)$$

$$\frac{dy_r}{dt} = -y_r + g'(H_r)\left\{\sum_i w_{ir} y_i + E_r\right\} \qquad (18.8)$$

يمكن أن ينظر للمعادلة (14.8) كعملية انتشار أمامي. خلال هذا الطور تسترخي الشبكة باستعمال هذه المعادلة لإيجاد القيم $x_i$ ومن ثم إيجاد الأخطاء الموافقة $E_i$.

توافق المعادلة (18.8) طور الإشارة في الانتشار الخلفي. تطبق هذه المعادلات عندما تسترخي الشبكة لإيجاد قيم $y_r$. أخيراً تستعمل المعادلة (17.8) لتنفيذ تحديث الأوزان.

يمكن استعمال الاستنتاجات السابقة، مع بعض التعديلات، لإيجاد خوارزمية تعليم الشبكات التكرارية التـــي تعمل بطريقة النقطة الثابتة. سنهمل التفاصيل هنا ونحيل القارئ المهتم إلى المراجع المذكورة سابقاً.

### 1.3.8 تدريب الشبكات التكرارية المتقطعة **Discrete Time RNN Training**

يمكن تدريب شبكات التكرار الديناميكية العامة غير المقيدة بشروط خاصة باستعمال خوارزمية الانتشار الخلفي التكرارية، حيث نُضمِّن بعد المسار الزمنـــي في حسابات الخطأ. وكما في الحالات السابقة للانتشار الخلفي، سنعرّف قياس الإنجاز أو تابع الكلفة كمتوسط مربع الخطأ، ولكن الآن عبر المسار الكامل للنماذج. وسنشتق خوارزميات تدريب الشبكة التكرارية العامة فيما يلي، لكن أولاً، سنقدم بعض التعاريف الأساسية.

ليكن الشعاع **x(t)** ببعد m يمثل المداخل الخارجية للشبكة في اللحظة t، وليكن **y(t)** يمثل مخارج كل وحدات الشبكة ببعد n في اللحظة t، وسنستعمل الشعاع **h(t)** ببعد q ليشير إلى الوحدات المخفية التـــي ليست وحدات دخل ولا خرج. من المفيد أيضاً تعريف الشعاع **z(t)** ببعد m + n الناتج عن **x(t)** و**y(t)** بالتناوب، ولتكن O مجموعة الأدلة k لوحدات الخرج في **z** و **I** مجموعة الأدلة لوحدات الدخل في **Z**. وستكون علاقة **z** مع **x** و **y** كما يلي:

$$z_k(t) = \begin{cases} y_k(t) & k \in O \\ x_k(t) & k \in I \end{cases} \qquad (19.8)$$

من الملاحظ أن **z(t)** لها بعض النسخ المتطابقة؛ أي إن كل وحدة دخل تظهر مرتين في **z**، مرة كمركبة **x** ومرة كمركبة **y**.

سنعتبر أيضاً قيمة $s_k(t)$ هي الدخل التركيبـــي المثقل net للوحدة k في اللحظة t. تحسب قيمة $s_k(t)$ في اللحظة t كتابع لحالة الشبكة ولدخل الشبكة عند اللحظة t كما يلي:

$$s_k(t) = \sum_{l \in O \cup I} w_{kl} z_l(t) \qquad (20.8)$$

يعطى خرج الوحدة k عند اللحظة t بـــ:

$$y_k(t+1) = f_k[s_k(t)] \qquad (21.8)$$

حيث $f_k$ تابع تفعيل الوحدة k. تعطي المعادلات (20.8) و(21.8) ديناميكيات الزمن المتقطع الكاملة للنظام.

يمكن أن تدرَّب الشبكات على نحوٍ متكيف أو غير متكيف. وإن جميع الخوارزميات الموصوفة من قبل كانت من النوع غير المتكيف. وقد دُرِّبت الشبكة عبر مجموعة ثابتة من أمثلة التدريب، وتغيرت الأوزان حتـــى تم الوصول إلى معيار ما، عند تلك النقطة أصبحت الأوزان ثابتة في كل العمليات اللاحقة. في حال تغير الوسط المحيط فيما بعد بطريقة ما، فإنه من الضروري إعادة تدريب الشبكة.

تدرب الشبكة، في حالة التدريب المتكيف، باستمرار خلال أطوار التدريب والعمل، أي لا يتوقف التدريب كما في حالة الشبكات المذكورة من قبل (غير المتكيفة). بالطبع يتطلب التدريب المتكيف، عند تغير الوسط المحيط، أن تكون بعض معطيات التدريب الجاري متوفرة خلال حياة النظام.

رأينا في الفصل الخامس أمثلة عن التدريب غير المتكيف في شبكات هوبفيلد والشبكات التكرارية الأخرى وذلك باستعمال تحميل حالات الذاكرة المحسوبة من قبل وتغيرات تعليم Hebb، ورأينا في الفصل الرابع والسادس بعض الأمثلة عن التدريب غير المتكيف أيضاً كقاعدة دلتا وقاعدة الانتشار الخلفي دلتا المعممة. وسنبحث الآن بكلتا الخوارزميتين المتكيفة وغير المتكيفة لشبكات التكرار العامة.

كما في حالة الشبكات العصبونية الصنعية المتعددة الطبقات الأمامية التغذية، سنُعرِّف قياس الإنجاز أو تابع الكلفة للشبكات التكرارية، حيث سنضيف في هذه الحالة البعد الزمنـــي للقياس. ليكن $d_k(t)$ الخرج المرغوب به أو القيمة المنشودة للوحدة k عند اللحظة t، وسنفترض وجود بعض قيم الخرج المنشود $d_k(t)$ كَمُعَلَّمٍ عبر المسار، وذلك على الرغم من أن القيم اللازمة لن تكون متوفرة عند كل زيادة خطوة زمنية.

لتمثيل هذه المجموعة من القيم سنستعمل T(t) كمجموعة من الأدلة $k \in O$ التـــي سيكون عندها قيم خرج منشود خاصة $d_k(t)$ بحيث تعطي الوحدة رقم k عند اللحظة t الاستجابة $y_k(t)$ لإشارة الدخل x(t)، وليكن $E_k(t)$ هو خطأ التغير الزمنـــي للوحدة رقم k عند اللحظة t المعطى بالعلاقة:

$$E_k(t)=\begin{cases} d_k(t)-y_k(t) & k\in T(t) \\ 0 & k\notin T(t) \end{cases} \qquad (22.8)$$

يسمح هذا التعريف بتخصيص قيم الخرج المنشود لوحدات مختلفة، وعند لحظات زمنية مختلفة إذا رغبنا في ذلك. وقد يكون هذا ضرورياً عندما لا تكون القيم المنشودة متوفرة عند كل لحظة تقطيع.

فمثلاً، فيما يتعلق بالخطأ الكلي عند الزمن t نضع:

$$E(t)=\frac{1}{2}\sum_{k\in O}\left[E_k(t)\right]^2 \qquad (23.8)$$

وسنهتم بتقليل هذا الخطأ عبر المجال الزمنـــي $[t_o,t_f]$، أي نريد تقليل $E_{tot}[t_o,t_f]$، حيث

$$E_{tot}(t_o,t_f)=\sum_{\tau=t_o+1}^{t_f}E(\tau) \qquad (24.8)$$

بعدئذ يعطى التعديل الكلي للأوزان $w_{ij}$ عبر المسار بالكامل بالعلاقة:

$$\Delta w_{ij}=\sum_{\tau=t_o+1}^{t_f}\Delta w_{ij}(\tau)=-\eta\sum_{\tau=t_o+1}^{t_f}\frac{\partial E(\tau)}{\partial w_{ij}} \qquad (25.8)$$

حيث $\eta$ معدل التعليم الموجب، وهذا التغير للأوزان سيكون متناسباً مع التدرج:

$$\Delta w_{ij}=-\eta\frac{\partial E(t)}{\partial w_{ij}}$$

$$-\frac{\partial E(t)}{\partial w_{ij}}=\sum_{l\in O}E_k(t)\frac{\partial y_k(t)}{\partial w_{ij}} \qquad (26.8)$$

لكن

$$\frac{\partial y_k(t+1)}{\partial w_{ij}}=f'_k[s_k(t)]\left[\sum_{l\in O}w_{kl}\frac{\partial y_1(t)}{\partial w_{ij}}+\delta_{ik}z_j(t)\right] \qquad (27.8)$$

حيث $\delta_{ik}$ هو دلتا Kronecker، أي $\delta_{ik}=1$ في حالة $i=k$ و $\delta_{ik}=0$ فيما عدا ذلك. ولما كانت الحالة الأولية للشبكة مستقلة عن الأوزان ، فإن:

$$\frac{\partial y_k(t_o)}{\partial w_{ij}}=0 \quad \text{لكل } K \qquad (28.8)$$

وبوضع

$$\frac{\partial y_k(t)}{\partial w_{ij}} = P_{ij}^k(t)$$

نستطيع تعريف المعادلات التكرارية التالية:

$$P_{ij}^k(t_o) = 0$$

$$P_{ij}^k(t+1) = f_k'[s_k(t)]\left[\sum_{l\in O} w_{kl}P_{ij}^k(t) + \delta_{ik}z_j(t)\right] \quad (29.8)$$

ومن ثم نعرف تحديث الأوزان الموافق بالعلاقة:

$$\Delta w_{ij}(t) = \eta \sum_{k\in O} E_k(t)P_{ij}^k(t) \quad (30.8)$$

نستطيع تخصيص خوارزميات متكيفة مختلفة لعملية التدريب بالاعتماد على مجموعة التدريب وعوامل أخرى. مثلاً، نستطيع تكديس تغيرات الوزن المحسوبة عند كل خطوة t على طول المسار الكلي باستعمال المعادلة (30.8) ومن ثم نعدل كل وزن بالمجموع (25.8).

وهذا يتطلب أن تكون المداخل وحالات الشبكة والأشعة المنشودة مخزنة عبر كامل المسار. يوفر مجموع الأخطاء عبر كامل المسار تحديث تدرج صحيح لتابع الخطأ. أيضاً، هذا يشبه معالجة كامل المسار كدور (طريقة المسار كدور). وبالتناوب يمكن أن تعدل الأوزان في الزمن الحقيقي عند كل خطوة زمنية على طول المسار، وذلك بتبسيط عملية التعديل وإزالة الحاجة لحدود الدور. في هذه الحالة، يجب أن تبقى خطوة التعليم صغيرة جداً لتقليل مقياس زمن تحديث الأوزان بالنسبة إلى عمل الشبكة. وهذا سيعطي تقريباً أفضل لتدرج صحيح يقلل تأثيرات أية تغذية عكسية سالبة وفقاً لتغيرات الوزن المنفذة على طول المسار.

من الواضح أن هذا التقريب الأخير يتطلب حساباً أكثر بكثير من طريقة المسار كدور. سندعو هذا التقريب الأخير بتعليم الانتشار الخلفي في الزمن الحقيقي RTBP (Real-Time BackPropagation learning).

فلحساب تحديث الأوزان من المناسب كتابة المعادلة (25.8) كما يلي:

$$\Delta w_{ij} = \eta \sum_{\tau=t_o+1}^{t_f} \delta_i(\tau)x_j(\tau+1) \quad (31.8)$$

حيث

$$\delta_k(\tau) = f'_k[s_k(\tau)]E_k(\tau) \quad , \quad if \;\; \tau = t_f \qquad (\text{أ} — 32.8)$$

$$\delta_k(\tau) = f'[s_k(\tau)]\left[E_k(\tau) + \sum_{l \in O} w_{lk}\delta_l(\tau+1)\right], if \;\; t_o < \tau < t_f \quad (\text{ب} — 32.8)$$

وهكذا يمكن تلخيص ما سبق كما يلي:

يمكن أن ينجز تعليم الانتشار الخلفي للدور خلال الزمن EBPT (Epochwise BackPropagation through Time learning) بترك الشبكة تجري خلال المجال الكلي $[t_O,t_f]$ وتخزين الدخل وحالة الشبكة وأشعة الخرج في كل لحظة $\tau \in [t_o, t_f]$، وبعدئذ أداء تمرير خلفي يبدأ عند اللحظة الزمنية الأخيرة $t_f$ لحساب $\delta_k(\tau)$ لكل $k \in O$ وكل لحظة $\tau \in [t_o, t_f]$ (المعادلة (32.8)).

بعدئذ تُستعمل المعادلات (31.8) و(25.8) لتحديث الأوزان في كل المسار (دور). يبدأ الحساب مع آخر خطوة زمنية ويزداد بالاتجاهات المعاكسة إلى الخطوات الزمنية السابقة من خلال تطبيق متكرر للمعادلة الثانية (30.8).

يمكن النظر إلى هذه الطريقة كطريقة انتشار خلفي عادي يطبق على شبكة متعددة الطبقات أمامية التغذية بأية قيم منشودة معطاة لعدد من الطبقات وليس قيماً منشودةً لطبقة الخرج فقط(كما في شبكات MLFF). بعد أن يكون الحساب قـــد نفذ عكسياً ووصل إلى $t_O + 1$، يمكن أن يفعل تغير الأوزان بما يوافق المعادلة (25.8) أي:

$$\Delta wij = \eta \sum_{\tau = t_o + 1}^{t_f} \delta_i(\tau) x_j(\tau + 1) = -\eta \frac{\partial E_{tot}(t_o, t_1)}{\partial w_{ij}} \qquad (33.8)$$

أما عندما تكون الشبكة قد حُدِّثت في الزمن الحقيقي عند كل لحظة t خلال عملية الشبكة RTBP فقط، فيصبح تخزين تاريخ المداخل وحالات الشبكة ضرورياً. عندئذ، يكون لكل t القيم:

$$\delta_k(\tau) = f'_k[s_k(\tau)]E_k(\tau) \quad ,if \;\; \tau = t \qquad (\text{أ} — 34.8)$$

$$\delta_k(\tau) = f'_k[s_k(\tau)]\left[E_k(\tau) + \sum_{l \in O} w_{lk}\delta_l(\tau+1)\right], if \;\; t_o < \tau < t \qquad (\text{ب} — 32.8)$$

المحسوبة في حالة $k \in O$ وكل $\tau \in [t_o, t_f]$، وتكون البداية عند أحدث خطوة زمنية t.

وبعدئذ ينفذ تعديل الأوزان حالاً بعد أن تجري حسابات الانتشار الخلفي عكسياً إلى $t_o+1$ باستعمال المعادلة:

$$\Delta w_{ij} = \eta \sum_{\tau=t_o+1}^{t} \delta_i(\tau) x_j(\tau-1) = -\eta \frac{\partial E_{tot}(t_o,t)}{\partial w_{ij}} \qquad (35.8)$$

لاحظ أنه في حالة RTBP لا يلزم أن تكون القيم المنشودة الأولى مخزنة لأن الخطأ فقط $E_k(\tau)$ في حالة $\tau = t$ يلزم في حساب المعادلة (34.8).

لسوء الحظ تحتاج هذه الخوارزمية إلى زمن حساب وتخزين ينمو خطياً مع زمن تنفيذ الشبكة، وهذا ليس مرغوباً به في المسائل العملية. نستطيع تحديد كمية الزمن والتخزين اللازمة بتقديم عدد ما من الخطوات الزمنية h كنافذة أو دور وتناسي كل شيء قبل الخطوات h التـــي هي أكثر حداثة. هذه هي خوارزمية الانتشار الخلفي المبتور خلال الزمن TBPT (Truncated BackPropagation through Time) وهي فقط تقريب للتدرج الصحيح.

يمكن أن تكون هذه الخوارزمية مناسبة عندما تعدل الأوزان خلال تنفيذ الشبكة. وفي هذه الحالة، من الضروري عند كل خطوة t حساب القيم $\delta_k(t)$ لكل $k \in O$ وفقط للأزمنة $\tau \in [t-h, t]$ باستعمال المعادلة (34.8). بعد حساب هذه القيم، يجري تحديث الأوزان باستعمال:

$$\Delta w_{ij} = \eta \sum_{\tau=t-h+1}^{t} \delta_i(\tau) x_j(\tau-1) \qquad (36.8)$$

من الواضح أن اختيار قيمة h صغيرة يؤدي إلى تقليل الحساب ومتطلبات التخزين، لكن عند حافة خطر الأداء (خطر الأداء الفقير). على أية حال، يمكن أن تكون قيمة صغيرة لـــ h كافية لبعض التطبيقات (التطبيقات مع تغير بطيء في الوسط المحيط).

أخيراً يمكن أن نطور خوارزمية تتطلب زمن حساب أقل مما هو مطلوب في حالة TBPT وهذا حل وسط بين EBPT وTBPT. يمكن أن تستعمل هذه الطريقة لقضايا يطبق فيها كل من هذه الخوارزميات. ويحدث هذا باختيار نافذة بعرض h كما سبق استخدامه للتعليم، ولكن الآن يمكن أن ينفذ التحديث بأقل تكرارية.

نختار قيمتين زمنيتين h و `h حيث `h≤h وننفذ حسابات $\delta_k(t)$ عبر المجال الأطـــول E[t–h, t] لكن بدلاً من إنجاز التمرير العكسي عند كل خطوة زمنية t، ينجز بعد تنفيذ الخطوات `h، والبداية عند الخطوة `t + h. هذا التقريب حل وسط يتطلب أن تكون المداخل وحالات الشبكة وقيم الخرج المنشودة مخزنةً في زمن (خلال) عمل الشبكة حيث لا تنجز أي عملية خلال الخطوات الزمنية `h . إن معادلات $\delta_k(t)$ في هذه الحالة هي:

$$\delta_k(\tau) = f_k'[s_k(\tau)]E_k(\tau) \quad , if\ \tau = t \qquad (أ — 37.8)$$

$$\delta_k(\tau) = f'[s_k(\tau)]\left[E_k(\tau) + \sum_{l \in O} w_{lk}\delta_l(\tau+1)\right], if\ t-h' < \tau < t \qquad (ب — 37.8)$$

$$\delta_k(\tau) = f'[s_k(\tau)]\left[\sum_{l \in O} w_{lk}\delta_l(\tau+1)\right] \quad ,if\ t-h < \tau \le t-h' \qquad (ج — 37.8)$$

وبعدئذ ينجز تحديث الأوزان باستعمال المعادلة:

$$\Delta w_{ij} = \eta \sum_{\tau=t-h+1}^{t} \delta_i(\tau) x_j(\tau-1) \qquad (38.8)$$

كما هو الحال في TBPT.

لاحظ أنه عندما يكون h`=1، فإن هذا الإجراء يؤول إلى TBPT، وعندما h`=h فإن الإجراء يكون تماماً EBPT. وكما نرى فإن اختيار قيم `h و h هام جداً. **إذا كانت النسبة `h/h صغيرة (قريبة من الواحد) فإن الإجراء سيكون أكثر فعالية. من ناحية أخرى، لتقريب أفضل لتدرج هبوط صحيح، يجب أن يكون الفرق `h - h كبيراً وهكذا فإن التسوية (أو الحل الوسط) يجب أن تفعل لتحقيق هذين المعيارين المتعارضين، هذا الإجراء الأخير يسمى** TBPT(h , h`)، وله مصاعب زمن وتخزين من رتبة $O(n^2h/h`)$ و O(n h) على الترتيب في شبكات التكرار الديناميكية بـــ n وحدة و $n^2$ وزناً.

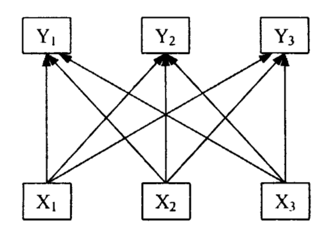
أظهرت تجارب المحاكاة أن طريقة TBPT(h,h`) تنجز أفضل من EBPT في عدة أنواع من المسائل لكن مع تقليل كافٍ لزمن الحسابات. مثلاً في مهمة توازن يمين يسار (مهمة تعليم Turing المنشورة من قبل Williams & Zipser عام 1989 [95])، كانت شبكة TBPT(16,8)بـــ 12 إلى 15 وحدة تجري 50 مرة أسرع من مثيلاتها EBPT. هناك محاولات

لتقصير زمن التعليم أنجزت من قبل باحثين آخرين. تتعلم شبكات التكرار الديناميكية عادة تصوير نفسها لحل مشكلة معطاة بطرق مختلفة بالاعتماد على خطة التدريب. مثلاً شبكة بطبقة ... نفس الشبكة ستصور كشبكة بثلاث طبقات عندما....

**مثال 1:**

**استعمال خوارزمية التدريب الخلفي في الزمن لتشكيل مسجل إزاحة بسيط.**

دربت الشبكة العصبونية بدون وحدات مخفية للعمل كمسجل إزاحة باستعمال خوارزمية الانتشار الخلفي في الزمن(Rumelhart , Hinton , Williams عام 1986 [56]])، سنعتبر مثلاً الشبكة الموضحة في الشكل (7.8) بثلاث وحدات دخل وثلاث وحدات خرج وكل وحدة لها مدخل انحياز. تتألف نماذج التدريب من كل الأشعة الثنائية بثلاث مركبات؛ وهذا يعني أن الهدف المنشود المرافق لكل نموذج دخل سيكون نموذجاً مزاحاً بمركبتين إلى اليسار (دوران دائري)، وهو يمثل الاستجابة المرغوب فيها للشبكة بعد خطوتين زمنيتين من المعالجة.

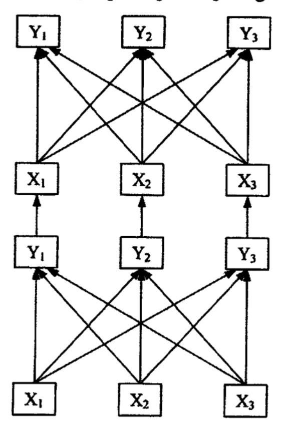



**الشكل 7.8**: ا لشبكة التكرارية المستعملة كمسجل إزاحة

النموذج الموسع للشبكة موضح في الشكل (8.8). يوضح هذا المثال حقيقة أنه ليس ضرورياً أن تكون لدينا معلومات عن الأخطاء عند الخطوات الزمنية المتوسطة.

إذا أعطيت الشبكة الاستجابة المنشودة بعد أول خطوة زمنية، سيكون الحل بسيطاً جداً. بدلاً من أن تكون الأوزان في كلتا نسختي الشبكة معدلة على أساس الأخطاء بعد خطوتين زمنيتين. على العموم، يمكن أن يستعمل تركيب المعلومات عن الأخطاء عند

المستوى النهائي وعند أي من المستويات المتوسطة أو كلها.




**الشكل 8.8:** المخطط الموسع لشبكة التكرار المستعملة كمسجل إزاحة

لقد وجد Rumelhart, Hinton, Williams عام 1986 -أ[100] و1986 -ب[101]، أن الشبكة المتدربة بقوة تطلبت، من أجل مسجل الإزاحة، 200 دور تدريب أو أقل بقليل، وبمعدل تعديل 0,25، وكانت أوزان الانحياز ملزمة بقيم سالبة دائماً. جرى التوصل إلى نفس النتائج مع شبكة بخمس وحدات دخل وخمس وحدات خرج. في غير هذه الحالات، إذا لم تفرض القيم السالبة على الانحيازات سنحصل على حلول أخرى للتدريب. حُصِلَ على هذه النتائج المرغوب فيها بعد عدد زوجي من الخطوات الزمنية وليس بعد عدد فردي من الخطوات الزمنية.

**مثال 2:**

**استعمال الانتشار الخلفي في الزمن لإعطاء تابع الجيب المتخامد.**

يمكن أن يستعمل الانتشار الخلفي في الزمن لتدريب الشبكة العصبونية لإعطاء تابع الجيب

المتخامد. تمثل وحدات الدخل قيم التابع بعدة خطوات زمنية، وتعطي وحدة الخرج قيمة التابع عند الخطوة الزمنية التالية. في هذا المثال البسيط، سيكون للشبكة أربع وحدات دخل، وخمس وحدات مخفية، كما هو موضح في الشكل (9.8). يعتمد عدد وحدات الدخل المتطلب على تردد الاهتزاز الجيبـــي ω في تابع الهدف المنشود:

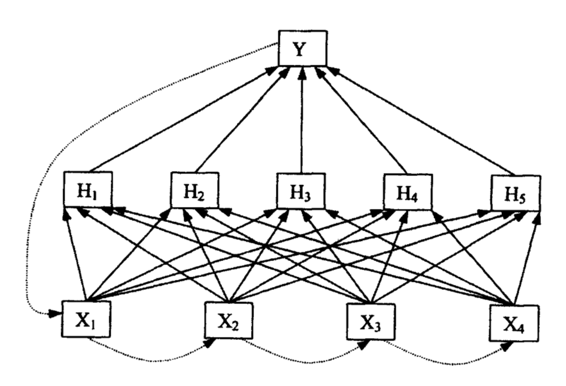
$$f(t) = \frac{\sin(\omega t)}{\omega t}$$

في حالة $w = \pi$ ، ستكون سبع وحدات دخل كافية. في اللحظة t، تستقبل الوحدة $X_1$ قيمة التابع f (t–1) المحسوب من الشبكة، من الوحدة Y، وتستقبل $X_2$ قيمة التـــابع السابقة f (t–2) من $X_1$، وتستقبل $X_3$ قيمة التابع f (t–3) من $X_2$ ، وتستقبل $X_4$ قيمة التابع f (t–4) من $X_3$. هنا ثانيةً، نستطيع التفكير بعملية التدريب للشبكة كأنها تتألف من عدة نسخ من الشبكة، لكن ليس من الضروري فعلياً برمجة كل نسخة على نحو منفصل. مثلاً فـــي حالة ω = 0.5 كان للشبكة 10 عقد دخل وعشر عقد مخفية وعقدة خرج واحدة.

ستكون عملية التدريب كما يلي:

1. أعط الأوزان القيم الأولية (قيماً صغيرة عشوائية).
2. مادام شرط توقف التدريب لم يتحقق، كرر الخطوات من 3 إلى 10.
3. أعط التفعيلات القيم الأولية( قيماً عشوائية صغيرة).
4. أعط قيمة التابع الأولية f (0) إلى وحدة الدخل $X_1$.
5. حتـــى تَحقُّق شرط توقف الأدوار، كرر الخطوات من 6 إلى 9.
6. احسب استجابة الشبكة : $y = f(1)$.
7. احسب الخطأ في الخطوة الزمنية الحالية،

 أوجد تحديث الأوزان بالانتشار الخلفي (لا تغير الأوزان)
8. حدّث التفعيلات: $x_4 = x_3, x_3 = x_2, x_2 = x_1, x_1 = y$
9. اختبر شرط توقف الدور: إذا كان y > max ، أو عدد الخطوات الزمنية أصبح أكثر من 30 عندئذ :طبق تحديث الوزان واستمر بالخطوة 10 وإلا استمر بالخطوة 4.




**الشكل 9.8:** شبكة عصبونية تكرارية لتنفيذ تابع الجيب المتخامد

10. اختبر شرط توقف التدريب:

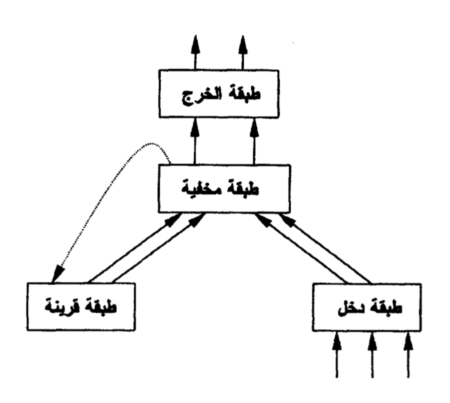
إذا كان الخطأ أصغر من مجال التسامح أو العدد الكلي للأدوار أصبح أكبر من حد التوقف عندئذ توقف، وإلا استمر بالخطوة 2.

## 4.8 بنـــى الشبكات العصبونية التكرارية البسيطة

## Simple Neural Network Architectures

قام بعض الباحثين أمثال Elman عام 1991 [102] وServan-Schreiber وزملاؤه عام 1991 [103] بتجارب على صنف من الشبكات العصبونية الصنعية ذات التغذية العكسية الجزئية فقط وأسموها الشبكات التكرارية البسيطة SRN (Simple Recurrent Networks).

في هذه الشبكات، تسمح مخارج الطبقة المخفية بتغذية عكسية لنفسها من خلال طبقة عزل (Buffer) أو قرينة (context). توفر هذه الطبقة فقط تغذية وصلات التغذية العكسية في الشبكة، ووضعت الأوزان على الوصلات بين الطبقة المخفية وطبقة القرينة بقيم ثابتة، أما جميع الوصلات الأخرى فكانت ذات تغذية أمامية بأوزان معدلة. يوضح الشكل (10.8) بنية شبكة تكرارية بسيطة نموذجية.




**الشكل 10.8**: الشبكة التكرارية البسيطة

وعلى الرغم من بساطة هذه البنية، فإنها قادرة على تعلم إنجاز مهام قوية.

لقد عولجت الإشارة في شبكة التكرار البسيطة في خطوتين زمنيتين. خلال الخطوة الأولى عند اللحظة t -1 تتوزع الإشارات من طبقة الدخل وطبقة القرينة التي تكون متصلة اتصالاً كاملاً مع الطبقة المخفية إلى وحدات الطبقة المخفية.

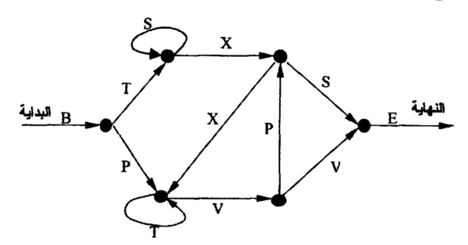
يحسب نموذج تفعيل مخارج الطبقة المخفية ويمرر إلى طبقة الخرج ليعالج عند اللحظة t، وبنفس الوقت تعاد مخارج الطبقة المخفية للخلف إلى مجموعة وحدات القرينة. تضم بعدئذ مخارج وحدات القرينة مع إشارات الدخل في الدورة التالية لتغذية الوحدات المخفية ثانيةً عند اللحظة t + 1، وهكذا تستمر العملية. تمزج المداخل الخارجية مع المداخل المحسوبة مسبقاً "أي مخارج طبقة القرينة" لتعطي اندماجات تكرارية للمداخل المحولة إلى طبقة الخرج عبر الطبقة المخفية.

تكون عادة أوزان وصلات التغذية العكسية من الطبقة المخفية إلى طبقة القرينة قيماً ثابتةً (غير قابلة للتعديل)، وتؤخذ على الأغلب واحدية، أما باقي الأوزان في الشبكة فتكون

قابلة للتعديل والتحديث. تُعَلَّم الأوزان المعدلة على ترميز متتاليات من نماذج الدخل خلال عملية التدريب. وتكون توابع التفعيل عادةً توابع تفاضلية غير خطية مع أنه في بعض التطبيقات تكون توابع تفعيل وحدات الخرج خطية.

لقد أثبتت الشبكات التكرارية البسيطة الموضحة في الشكل (10.8) إمكانية القيام بحسابات منتهية مكافئة لجهاز ذاتـــي الحركة وذلك بسلسلة من التجارب التـــي نفذت على شبكات تكرارية بسيطة مُعلَّمة على ترميز تتابع مسافة طويلة في سلاسل حرفية متولدة مـــن قـــاعدة الحالة المنتهية FSG (Finite-State Grammar) التـــي تسمـــى قاعدة Reber (Servan-Schreiber عام 1991 [103]).

السلاسل الحرفية القواعدية هي سلاسل متولدة بواسطة مخطط عبور قواعدي منتهي الحالة المبين في الشكل (11.8).



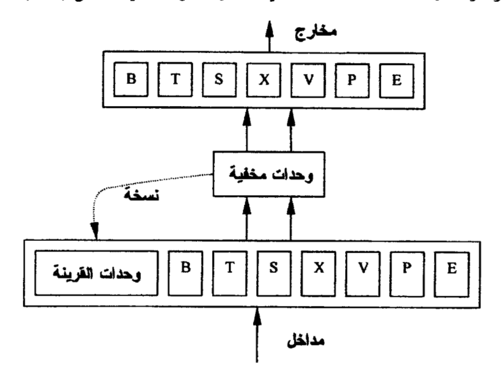

**الشكل 11.8**: مخطط العبور لقاعدة الحالة المنتهية لـــ Reber.

يبدأ العبور عند العقدة B اليسرى من المخطط، وتتولد السلسلة الحرفية بواسطة الانتقال عبـــر المخطط باتجاه الأسهم من عقدة إلى أخرى، والاختيار لخط المـــرور عند كـــل عقدة (تفريعة سوداء) باحتمال متساوٍ؛ أي إن مجموع إمكانيات الممرات الصادرة عن العقدة يساوي الواحد. تحدد الوصلة على الممر المختار الحرف التالي في السلسلة والعقدة التالية في ممر العبور من اليسار إلى اليمين.

تتكرر عملية الاختيار الاحتمالية للممرات عند كل عقدة حتـــى يتم الوصول إلى العقدة

E النهائية. إن الكلمة أو السلسلة الحرفية المتولدة من خلال العبور التام للمخطط هي واحدة من السلاسل الصحيحة لقاعدة الحالة المنتهية. وبسبب حلقتـــي التكرار سيكون لدينا عدد غير محدود من السلاسل المتولدة، مثل السلسلتين الصحيحتين التاليتين بطولين مختلفين : BPVVE وBTSSSXXTVPSE. صممت التجارب لإثبات أن شبكة التكرار الديناميكية قادرة على تمييز صحيح للسلاسل الحرفية المتولدة من قاعدة Reber.

استُعملت لتنفيذ التجارب شبكة تكرارية مؤلفة من 7 وحدات دخل و7 وحدات خرج (وحدة واحدة لكل حرف من الأحرف السبعة المستعملة في المخطط بما في ذلك خطوط البداية والنهاية، مع الأشارة إلى أن ترتيب وحدات الأحرف كيفي)، و3 عقد في الطبقة المخفية و3 عقد في طبقة القرينة. ونفّذت أيضاً تجارب مع عدد ضخم من العقد في الطبقة المخفية والقرينة. إن بنية الشبكة المستعملة في التجارب موضحة في الشكل (12.8).




**الشكل12.8**: شبكة التكرار المستعملة لتعلم السلاسل الحرفية المشكلة من قاعدة Reber

قدم للشبكة كدخل في كل محاولة تجريبية حرف ينتمي إلى السلسلة الحرفية، ومن المتوقع والمأمول أن تتنبأ الشبكة بالحرف التالي في السلسلة عند طبقة الخرج. استعمل لتدريب الشبكة مجموعة مؤلفة من 60000 سلسلة حرفية. ولدت السلاسل عشوائياً باستعمال قاعدة

Reber حيث اختير كل حرف في السلسلة بواسطة اجتياز مخطط العبور واختيار واحد من قوسين ممكنين متتابعين من كل عقدة ، كل قوس باحتمال 0,5، وكل سلسلة تبدأ بـــالحرف B وتنتهي بـــالحرف E. رتبت السلاسل في مجموعة التدريب وفقاً للحجم من 5 إلى 23 حرفاً مع حجم متوسط بسبعة أحرف (باستثناء B وE).

خلال التدريب قدم كل حرف إلى الشبكة كشعاع ثنائي ذي بعد 7. وركبت الأحرف سلسلة قدمت على التتالي، وكما ذكر فإن البداية عند B والنهاية عند E. بعد تقديم كل حرف جرى حساب الخطأ بين تنبؤ الشبكة والتتابع الفعلي في السلسلة وانتشار هذا الخطأ إلى الخلف باستعمال خوارزمية تدريب الانتشار الخلفي المناسبة. وُوضع تفعيل عقد طبقة القرينة بقيمة الصفر عند بداية كل حرف سلسلة، وذلك لإزالة أي قيمة تفعيل متبقية من السلسلة السابقة. استمر تنفيذ التدريب حتـــى وصل الخطأ المحسوب إلى مستوى مقبول جداً. وهذا يحدث بعد 2000 حتـــى 10000 دور تدريب. وقد استعمل لاختبار الشبكة مجموعتان مختلفتان من السلاسل المتولدة عشوائياً.

المجموعة الأولى مؤلفة من 20000 سلسلة ولدت مباشرة من القاعدة. وقد اعتبر جواب الشبكة صحيحاً إذا استطاعت الشبكة التنبؤ بكل حرف تالٍ للحرف المقدم كدخل في السلسلة المعطاة. واعتبر التنبؤ صحيحاً إذا كانت قيمة تفعيل عقدة الخرج أكبر من 0.3 (الموافق لاحتمال حوالي 0.5). إذا لم يتحقق هذا المعيار، يوقف تمثيل السلسلة وتعتبر مرفوضة. بهذا الاختبار كان تنفيذ الشبكة تاماً جداً، وقد تنبأت على نحوٍ صحيح بكل متتاليات الأحرف في 20000 سلسلة مستعملة في مجموعة الاختبار.

وتألفت مجموعة الاختبار الثانية من 30000 سلسلة متولدة عشوائياً وكان معظمها غير مبنـــي لغوياً، فقط 0,2 من هذه المجموعة (26 سلسلة) كان مبنيّاً لغوياً؛ السلسلة غير اللغوية كانت مشكّلة من نفس الأحرف المستعملة في قاعدة Reber، ولكنها كانت ليست صحيحة (عبور خاطئ للمخطط). وهنا أيضاً أنجزت الشبكة على نحوٍ صحيح جداً، حيث رفضت كل السلاسل غير القواعدية وقبلت فقط السلاسل القواعدية الصحيحة، حتـــى عندما قدم إلى الشبكة كدخل سلسلة طويلة جداً (مشكلة من 100حرف متتالي) كما يلي:

"BTSSSSSSSSSSSSSSSSSSSSSSXXVPXVPXVPXVPXVPXVPXVPXPXV
PXVPXVPXTTTTTTTTTTTTTTTTTTTTTTTTTVPXVPXVPXVPXVPXVPX"

فقد تنبأت على نحوٍ صحيح بالحرف التالي وليس بآخر.

لفعل ذلك، كان للشبكة تمثيلات تعليم في الطبقة المخفية، حيث نُسخت هذه التمثيلات عكسياً إلى طبقة القرينة التـــي رمَّزت المكان الحالي لقاعدة الدخل التـــي تدربت الشبكة عليها. وقد أجري تحليل تجمعّ على تفعيلات عقد الطبقة المخفية خلال الاختبار ساعد على إظهار بُنية شجرة تصاعدية طُورت في التمثيلات.

جُمِّعت نماذج التفعيل بمجموعات وفقاً للعقد المختلفة في قاعدة الحالة المنتهية. وجُمِّعت أيضاً بمجموعات النماذج التـــي أعطت تنبؤات متشابهة.

نفذت تجارب معقدة أكثر في بعض الاختبارات، استُخلص منها أن شبكة التكرار تستطيع أن تتعلم لتسلك كجهاز ذاتـــي الحركة المنتهي مع إمكانية تعلم قاعدة الحالة المنتهية والتعميم على النماذج المعلمة.

أُجريت بعض التعديلات على الشبكات التكرارية البسيطة السابقة وذلك بإضافة طبقات مخفية إضافية، وتغيير وصلات طبقات القرينة. مثلاً، يمكن أن تضاف طبقة ثانية مخفية بوصلات خرج يمكن أن تنسخ عكسياً للطبقة المخفية الأولى أو عكسياً إلى نفس الطبقة كما هو موضح في الشكل (13.8).

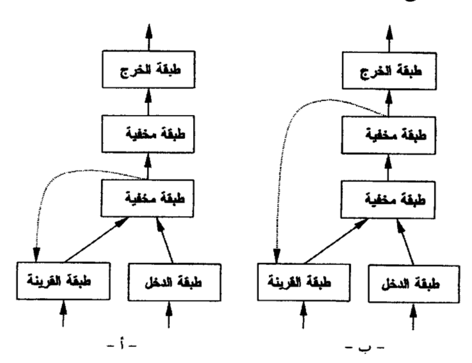
اختبرت هذه الشبكات اختباراً موسعاً في تطبيقات التنبؤ المالي وسلاسل الزمن الفوضوية من قبل Patterson عام 1993 [3]، والكثيرين مثل Sterzing & Schurmann 1993 [104]، وLiang & haykin عام 1993 [105]، وMori & Ogasarawa عـــام 1993 [106]، وRao & Ramamurti عام 1993 [107]، وقد لا نستطيع ذكرها وعلى القارئ مراجعة هذه المصادر.

هناك شبكات تكرار أخرى هجينية تستعمل مزيجاً من توابع التفعيل **sigmoid** وتوابع الأساس الشعاعي بتغذية عكسية جزئية استعملت في تنبؤ سلاسل الزمن الفوضوية. وسندرس هذه الشبكات وأخواتها الأخريات فيما بعد.

**مثال 3:**

لتكن لدينا السلسلة الحرفية: B T X S E المتولدة من قاعدة **Reber**، وسنحاول كتابة خوارزمية تدريب الشبكة التكرارية الديناميكية. ستقدَّم الأحرف إلى دخل الشبكة حرفاً تلو

الآخر بشكل شعاع مؤلف من ستة مركبات (مركبة لكل حرف من الأحرف الخمسة، والمركبة السادسة لحرف البداية)، مثلاً الحرف B سيوافق الشعاع (0 ,0 ,0 ,0 ,0 ,1). عند بداية التدريب توضع تفعيلات وحدات القرينة بقيمة 0.5.




**الشكل 13.8**: إضافة طبقة مخفية ثانية لشبكة التكرار البسيطة،

-أ- دخل طبقة القرينة من الطبقة المخفية الأولى،

-ب- دخل طبقة القرينة من الطبقة المخفية الثانية.

يوضح الشكل (14.8) بنية هذه الشبكة التكرارية التـــي لها 6 مداخل خارجية و3 عقد قرينة و3 عقد مخفية و6 مخارج (نود التذكير بأن ترتيب وحدات الأحرف في طبقتـــي الدخل والخرج كيفي، حيث يجري وفق هذا الترتيب تكوين الأشعة الثنائية لإدخال كل حرف من السلسلة الحرفية وإخراجه).

في البداية سنستعرض خطوات الخوارزمية عموماً، وستكون على النحو التالي :

في كل سلسلة حرفية للتدريب كرر الخطوات من 1 إلى 7.

1. ضع تفعيلات وحدات القرينة بقيمة 0.5.

2. كرر الخطوات من 7.3 حتـــى نهاية سلسلة التدريب الحرفية:

3. قدم الحرف الأول كدخل.
4. قدم الحرف التالي لوحدات الخرج كاستجابة خرج منشود.
5. احسب الحرف التالي المتنبأ به.
6. حدد الخطأ واستعمل الانتشار الخلفي لتعديل الأوزان.
7. افحص شرط التوقف: إذا كان الهدف المنشود يساوي E عندئذ توقف وإلا انسخ فعيلات الوحدات المخفية إلى وحدات القرينة وتابع تنفيذ الخوارزمية.

في المثال المعطى لدينا: B T X S E ستكون خطوات الخوارزمية كما يلي:

2. بداية التدريب لهذه السلسلة
   3. الحرف الأول B كدخل؛ أي الدخل سيكون الشعاع (1, 0, 0, 0, 0, 0)
   4. الاستجابة المنشودة هي الحرف T؛ أي الشعاع (0, 0, 0, 1, 0, 0)
   5. احسب الاستجابة المتنبأ بها؛ شعاع بقيمة حقيقية بمركبات بين الصفر والواحد
   6. حدد الخطأ، واستعمل الانتشار الخلفي لتحديث الأوزان
   7. انسخ تفعيلات الوحدات المخفية إلى وحدات القرينة
2. التدريب للحرف الثانـــي في السلسلة
   3. الحرف الثانـــي T كدخل؛ أي الدخل سيكون الشعاع (0, 0, 0, 1, 0, 0 )
   4. الاستجابة المنشودة هي الحرف X؛ أي الشعاع (0, 0, 0, 0, 0, 1)
   5. احسب الاستجابة المتنبأ بها؛ شعاع بقيمة حقيقية بمركبات بين الصفر والواحد
   6. حدد الخطأ، واستعمل الانتشار الخلفي لتحديث الأوزان
   7. انسخ تفعيلات الوحدات المخفية إلى وحدات القرينة
2. التدريب للحرف الثالث في السلسلة
   3. الحرف الثالث X كدخل؛ أي الدخل سيكون الشعاع (0, 0, 0, 0, 0, 1 )
   4. الاستجابة المنشودة هي الحرف S؛ أي الشعاع (0, 0, 1, 0, 0, 0)
   5. احسب الاستجابة المتنبأ بها؛ شعاع بقيمة حقيقية بمركبات بين الصفر والواحد
   6. حدد الخطأ، واستعمل الانتشار الخلفي لتحديث الأوزان
   7. انسخ تفعيلات الوحدات المخفية إلى وحدات القرينة

2. التدريب للحرف الرابع في السلسلة

4. الحرف الرابع S كدخل؛ أي الدخل سيكون الشعاع (0, 0, 1, 0, 0, 0)

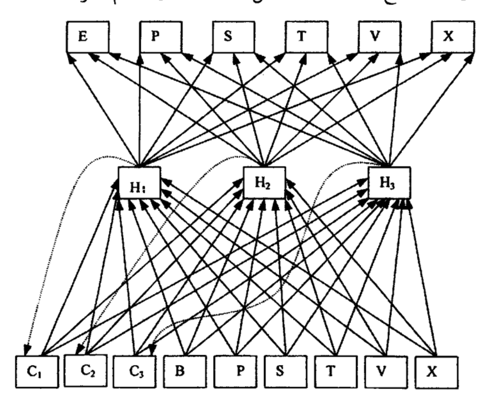
5. الاستجابة المنشودة هي الحرف E؛ أي الشعاع (1, 0, 0, 0, 0, 0)

6. احسب الاستجابة المتنبأ بها ؛ شعاع بقيمة حقيقية بمركبات بين الصفر والواحد

7. حدد الخطأ ، واستعمل الانتشار الخلفي لتحديث الأوزان

3. الاستجابة المنشودة هي حرف النهاية؛ لقد تم التدريب على هذه السلسلة.

بعد التدريب تستطيع السلسلة تحديد: هل السلسلة صحيحة؟ أم لا وفقاً لقاعدة Reber.




**الشكل 14.8**: شبكة تكرارية بسيطة لتعلم قاعدة الحالة المنتهية

## 5.8 تطبيقات الشبكات التكرارية

## Applications of Recurrent Networks

إن الشبكات التكرارية من النوع الموصوف في هذا الفصل ذات تطبيقات عملية محدودة في مسائل العمل الحقيقية. وهذا مفهوم طبعاً لأن هذه الشبكات ليست معروفة كثيراً وسلوكها ليس مفهوماً جيداً، وما تزال في الوقت الحالي تحت مجهر البحث العلمي.

بدأت تظهر حديثاً عدة مقالات علمية عن التطبيقات العملية للشبكات التكرارية في مختلف الميادين، وهذا ما سيجعل لهذه الشبكات تطبيقات عملية أكثر فأكثر، ومعقدة في الزمن الحقيقي، ومن ثم سيلزم هذه المسائل العملية تطبيقات (mappings) زمنية مكانية تحتاج إلى قدرة على حساب أنظمة التغذية العكسية المترابطة. ومع أن التطبيقات ما تزال محدودة، فقد حصل على بعض النتائج الفعالة في بعض الحقول كالتحكم، وتحقيق الشروط المقيدة، وتمييز أحرف الكتابة اليدوية وإشارة الكلام، والرؤية، والتنبؤ. فيما تبقى من هذا الفصل سنبحث في بعض هذه التطبيقات المستعملة لشبكات التكرار الديناميكية.

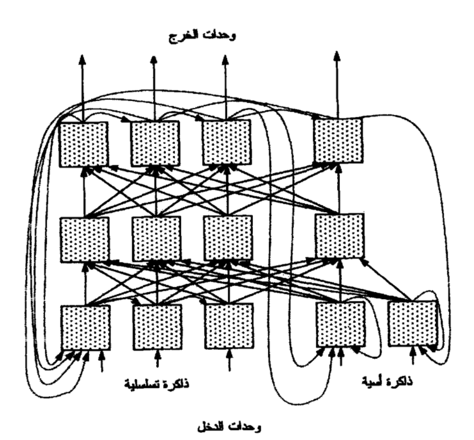
### 1.5.8 تركيب ألحان صوتية متعددة

**Composition of polyphonic Melodies**

درِّبت الشبكات العصبونية الصنعية التكرارية على تعلم مُلحِّن موسيقي في محاولة لتشكيل موسيقى جديدة، ولكن بأسلوب مشابه للملحن. واحدٌ من أكثر المشاريع طموحاً نشره Freisleben عام 1992 [108]، يَستعمل شبكة تكرارية لتأليف ألحان صوتية متعددة؛ أي تتألف من طبقة صوتية لحظية متعددة.

درِّبت هذه الشبكات على ستة ألحان ثنائية صوتية لـــ 100 مقطوعة زمنية لكل منها. مجموعتان من معطيات التدريب كانتا: غناء شعبـــي ألمانـــي وعازف كمان ثنائي لموزارت. وكانت البنية المستعملة في هذا التطبيق هي شبكة التكرار الموسعة المذكورة في الفقرة (4.8)، مع خطوط تغذية عكسية كثيرة. استعملت مجموعتان من وحدات الدخل؛ كذاكرة تسلسلية وذاكرة أسية.

صممت الذاكرة الأسية لتنسى تدريجياً قيمتها السابقة. واتصلت وحدات الدخل اتصالاً كاملاً بوحدات الطبقة المخفية، التـــي كانت بدورها متصلة اتصالاً كاملاً أيضاً بوحدات الخـــرج. ووُصلت وحدات الخرج عكسياً بذاكرتـــي الـــدخل كما هو موضـــح في الشكل (15.8). لاحظ أن الذاكرة التسلسلية لها وصلات تغذية عكسية متعددة بالوحدة الأقصى يساراً وخرج هذه الوحدة وصل جانبياً بوحدة الذاكرة التسلسلية المجاورة التـــي بدورها متصلة بمجاوراتها وهكذا. أما الوحدة الأخيرة في هذه الذاكرة فليس لها وصلات جانبية.




**الشكل 15.8**: شبكة التكرار الموسعة

لكل من وحدات الذاكرة الأسية وصلات تغذية عكسية مباشرة من وحدة الخرج بالإضافة إلى وصلات تغذية عكسية ذاتية. بهذا الترتيب لوصلات الذاكرة الأسية تعمل كذاكرة دور طويل لتمثيل لحن كامل، على حين تخزن الذاكرة التسلسلية بعض النغمات الأخيرة فقط. إن وحدات الخرج موصلة جانبياً لتدخل في منافسة "الرابح يأخذ الكل". عدد عقد الخرج محدد بواسطة عدد القطع المختلفة ضمن مجال الطبقة الصوتية المعتبرة في الألحان.

في تجارب للتدريب كان لوحدات الطبقة المخفية أوزانٌ قابلة للتعديل؛ ستعدل خلال الانتشار الخلفي، وكانت الأوزان على خطوط التغذية العكسية ثابتة.

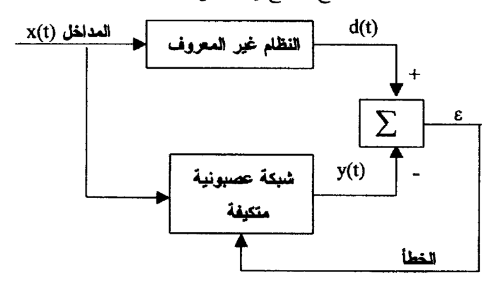
لقد استعملت في الشبكة 50 وحدة قرينة (ذاكرة دخل)، 25 لكل ذاكرة، ومع أنها لم تظهر في الشكل فقد كانت هناك شبكات متعددة متصلة من خلال وصلات التغذية العكسية، حيث دربت كل شبكة على كل صوت، واستُعمل حتـــى 50 وحدة طبقة مخفية لكل وحدة صوت، و25 وحدة خرج لكل صوت. علّمت الشبكة على ألحان فردية كاملة، وكانت قادرة على متابعتها في أسلوب نمطي مشابه للأصل.

### 2.5.8 تطبيقات التحكم Control Applications

تتطلب مسائل أنظمة التحكم عادة تطبيقات (mapping) زمنية غير خطية لإشارات الدخل. غالباً ما تكون ديناميكيات هذه الأنظمة معروفة، لذا يبدو أن شبكات التكرار الديناميكية يمكن أن تكون مرشحة لمهام التحكم إذا كانت معطيات تدريب الدخل والخرج متوفرة للنظام. هذا وتبدو التطبيقات الهامة للشبكات التكرارية في مجال التحكم ضخمةً جداً.

يمكن أن يعبر عن مسألة التحكم العام كما يلي: لدينا نظام معطى بديناميكيات غير معروفة، لكي نكوّن عنصرَ تحكمٍ مناسباً للنظام علينا البحث عن نموذج لهذا النظام؛ النموذج (model) هو أية وسيلة أو أداة تقلد سلوك النظام.

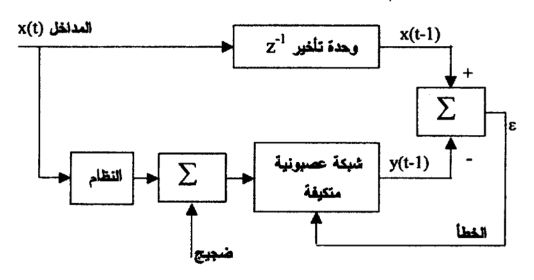
تسمى عملية إنشاء النموذج عندما تكون العلاقة بين مداخل ومخارج النظام متوفرة بتعريف النموذج(Identification)، ويسمى النموذج نفسه بالمُعرِّف (Identifier). حالاً بعد توفر النموذج، يمكن أن يكوِّن النموذج المعاكس ليخدم كعنصر تحكم (controller) في النظام الحقيقي. يحدث هذا النوع من المسائل في الكثير من أنظمة التحكم بما في ذلك أنظمة الربوت، ومحركات القيادة لأنظمة متنوعة (أنظمة تحميل وتفريغ الروافع)، والتحكم باللحام الآلي، وهكذا. سنصف فيما يلي مثالين اثنين من هذه التطبيقات. مبدأ استعمال الشبكات العصبونية الصنعية المتكيفة كنموذج موضح في الشكل (16.8).




**الشكل 16.8**: تعليم شبكة عصبونية متكيفة على نمذجة نظام غير معروف

تعدَّل أوزان الشبكة العصبونية الصنعية التـــي تستقبل نفس إشارة دخل النظام غير

المعروف على نحوٍ متكيف حتـــى يضبط خرجها ويكون قريباً من خرج النظام. بعد أن تصبح الشبكة العصبونية مدربة على نموذج النظام تبدأ عملية إنشاء عكس النموذج. يكون بعدئذ النموذج العكسي قادراً على العمل كعنصر تحكم بالنظام. يوضح الشكل (17.8) نظاماً نموذجياً لعنصر تحكم .




**الشكل 17.8**: عنصر تحكم شبكة عصبونية متكيف لنظام

حيث أضيفت وحدة التأخيـــر الزمنية لملاحظة تأخيــــر الإشارة خلال حلقة عنصر التحكم/النظام ، وأضيف الضجيج في الحلقة لتمثيل النظام الحقيقي.

لاحظ أن الشبكة العصبونية تحاول قيادة النظام لإعطاء الخرج الذي يضبط الدخل المتأخر ليعطي خطأً يساوي الصفر. بفعل هذا، يجب أن تتعلم عكس ديناميكيات النظام. وبهذه الملاحظات المقدمة، أصبحنا قادرين على البحث في تطبيق خاص للشبكة العصبونية التكرارية في نظام التحكم في رافعة جسرية متحركة وإليك التفاصيل.

### 1.2.5.8 عنصر تحكم في رافعة جسرية متحركة Gantry Crone Controller

الرافعة الجسرية أداة رفع إلكتروميكانيكية مستعملة في المصانع لتحريك الأجزاء الثقيلة من مكان إلى آخر. تتحرك هذه الرافعة من نقطة إلى نقطة أخرى في البنية الرأسية للمصنع، وتستعمل مجموعة كِبال قابلة للرفع لجعل الحمل مرفوعاً ومتحركاً. يشغل نظام التحكم بالرافعة محرك قيادة يعطي حركة أفقية للرافعة والحمل.

ستعمل ميكانيكية التحكم بحيث تتحرك الرافعة إلى موقع جديد مخصص بإحداثيات المكان، على حين تبقى حركة الحمل مثبطة لمنع حركة شاذة أو اهتزاز، مع السماح بوجود كتلة حمل متغيرة وأطوال كبل إضافية. ستحافظ وحدة التحكم على درجة عالية من الاستقرارية.

يَستعمل النظام حساسات للمكان والسرعة لمراقبة الحركة، ويتطلب ممر تحكم حلقة مغلقة مع عرض حزمة كبير يسمح بحمل متغير وأطوال كبل إضافية.

يمكن أن يكون النظام ممثلاً بمجموعة من المعادلات التفاضلية غير الخطية في وصف متحولات مكان الرافعة، والزاوية، وطول الكبل، وكتلة الرافعة، وثقل الحمل، وعوامل التثبيت، والقوة المطبقة. ويمكن أن تعاد كتابة المعادلات التفاضلية للحصول على تعابير الفروق الزمنية في حدود سرعات الرافعة، والحمل عند الزمن t +1 كتوابع للجهد الكهربائي المطبق على المحرك، والسرعات عند الزمن t، ووسطاء أخرى. ومن المتوقع تحقيق بقية الوسطاء للنظام بواسطة عنصر تحكم شبكة عصبونية.

في مسألة التحكم برافعة، درِّبت شبكة تكرار ديناميكية وشبكة متعددة الطبقات أمامية التغذية MLFF لتنفيذ تعريف النظام وتعريف النظام المعاكس من قبل Fernando عام 1992 [109]. تتطلب مهمة تعريف النظام أن تتعلّم الشبكةُ محاكاةَ النظام، وتتطلب مهمة تعريف النظام المعاكس أن تنفِّذ الشبكة وظيفةَ عنصر التحكم.

لقد ولِّدت 240 نقطة معطيات للتدريب والاختبار مع استعمال أعداد متساوية من النقاط لكل قيمة. وولِّدت مجموعة المعطيات من معادلات تفاضلية تصف النظام. وأُجريت اختيارات العشوائية لمجال الجهد من 0-200 فولت عند كل خطوة زمنية، ثم جرى حساب السرعات الموافقة. إضافة إلى ذلك، استُعملت الجهود الجيبية لمعطيات الاختبار لمقارنة النتائج في وضعيات أكثر تحققاً.

في عملية تعريف النظام، استُعملت شبكة تكرارية بإشارة دخل للجهد ومخرجين، واحد لكل من سرعات الرافعة والكبل، وكان للشبكة خمس عقد مخفية. ودُربت شبكة متعددة الطبقات أمامية التغذية MLFF لهذه المهمة كشبكة تأخير زمني. وكان الدخل التركيبي المثقل لهذه الشبكة هو خمس عقد دخل، واحدة للجهد وأربع لسرعات الرافعة والحمل؛

اثنتان عند الزمن t واثنتان عند الزمن t–1. وكانت مخارج الشبكة سرعتين عند الزمن t+1، واستُعملت أربع عقد مخفية. وُنفِّذ اختبار البنـــى لكلا الشبكتين بعد تنفيذ تجارب عديدة.

ولتعريف النموذج العكسي، استُعملت نفس مجموعة المعطيات التـــي تتألف من 240 نقطة للتدريب والاختبار. يوجد لشبكة التكرار المستعملة لهذه المسألة عقدتا دخل لسرعات الرافعة والحمل عند الزمن t، وثلاث عقد طبقة مخفية وعقدة خرج واحدة لجهد المحرك عند الزمن t–1.

استُعملت شبكة متعددة الطبقات أمامية التغذية MLFF لنفس المهمة، ولهذه الشبكة ست عقد دخل للسرعتين عند الأزمنة t وt–1 وt–2. واستُعملت عقدتان في الطبقة المخفية وعقدة خرج واحدة لجهد محرك التحكم عند الزمن t–1. وقد تعلمت كلتا الشبكتين مهمة تعريف النظام جيداً.

كان الخطأ على معطيات التدريب أصغر من الخطأ على شبكة متعددة الطبقات أمامية التغذية (متوسط مربع الخطأ الذي كان 0,005 أصبح 0,0003 لشبكات التكرار)، لكن إنجاز شبكات التكرار **كان أعلى قليلاً** على معطيات الاختبار.

على أية حال، وكما ذكر من قبل، تبدي شبكات التكرار حسنات أخرى؛ فهي أولاً لا تتطلب معرفة سابقة عن البنية الزمنية للنظام، وهذا معاكس تماماً للشبكة المتعددة الطبقات الأمامية التغذية، وثانياً تتطلب هذه الشبكة الأخيرة MLFF إبرازاً واضحاً لنقاط المعطيات الماضية للتدريب، على حين لا تتطلب شبكات التكرار ذلك.

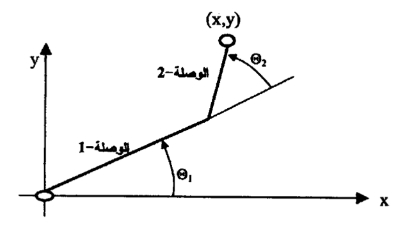
### 2.2.5.8 التحكم بذراع معالج Manipulator Arm Control

ثمة حقل خصب لتطبيقات التحكم في الربوت. ويمكن أن تكون مهمات التحكم بالربوت تحدياً كبيراً لأن فراغ حالة النظام الذي يجب أن يوظف قد يكون ضخماً جداً. ومع ذلك فقد أثبتت الشبكات العصبونية الصنعية أنها اختيارات منافسة لوحدات التحكم في هذا الحقل بسبب مقدرة التعليم المتكيف وإمكانية التعميم. لتوضيح المبادئ الأساسية سنصف مسألة التحكم بذراع ثنائي الوصلة الموضح في الشكل (18.7).

لتصميم عنصر تحكم في المعالج يمكن استعمال نفس التقريب المستخدم في حالة الرافعة

الجسرية المشروحة سابقاً. وكما ذكر من قبل، نحتاج أولاً إلى تعريف المعالج، وبعدئذ تكوين عنصر التحكم لهذا المعالج. على أية حال ليس من السهل تعريف نموذج المعالج، لأنه من غير الممكن استعمال كل المعطيات في فراغ الطور؛ وهناك مسارات تدريب متعددة ممكنة للتوقع.

لقد رشحت طريقة الاختصار المقدمة من قبل Hoshino عام 1991 [110] لتدريب الشبكة. فهي تَستعمل معادلات تعريف الحركة التـــي تكون معروفة. ويمكن أن تعطي معادلات النموذج تقريباً لمعطيات الحالة التـــي يمكن أن تُستعمل لتدريب الشبكة التكرارية تدريباً فعالاً. لهذا يلزم وجود وحدة إضافية (سمِّيت المعرف k) لتوفير معطيات المسار. إن مُعرِّف النظام المعدل موضح في الشكل (19.8).



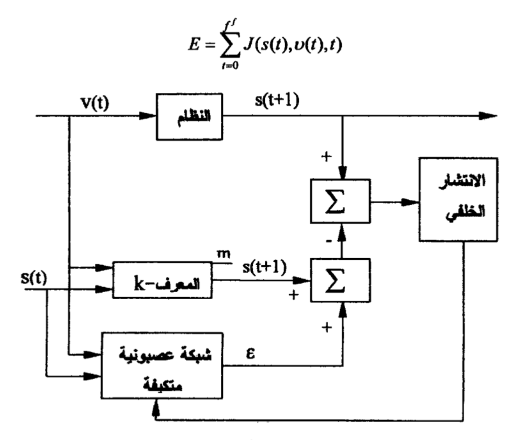

**الشكل 18.8**: ذراع معالج ثنائي الوصلة

في هذا الشكل: v(t) هو شعاع التحكم عند الزمن t، وs(t) هو شعاع الحالة. إن v(t) تابع لحالة s(t) وهو يتبع بدوره لقيم زاوية الوصلة ومقادير التغير في قيم الزاوية.

استُعمل الانتشار الخلفي لتدريب معرِّف نموذج الشبكة التكرارية المبنـــي على خطأ الفرق بين حالة النظام عند t + 1، وهي(s(t+1، وخرج شبكة المعرِّف k عند الزمن t + 1. يتعلّم المعرف العصبونـــي بخرج ε الفرق كإشارة تدريب.

إن التحكم في المعالج يتوقف على مكان الهدف المنشود $t^f$ ضمن مدة الزمن المسموح بها. وهذا ما يعبر عنه كقيد للإنجاز؛ وهو التابع E لتابع التحدب J الذي يعتمد على الحالة، والسرعة، وعزم دوران الذراع، ويعطى بـــ:

$$E = \sum_{t=0}^{t^f} J(s(t), \upsilon(t), t)$$




**الشكل19.8:** نظام معرّف معدل للمعالج

استُعملت شبكة تكرار عصبونية بطبقة مخفية واحدة للمحاكاة، وكان لشبكة المعرف 20 عقدة وعشر شبكات عنصر تحكم. وضعت قيم الزوايا في بداية التعليم مساوية للصفر، وعولجت الحالات المتتابعة بخطى زمنية مقدارها 0.02 ثانية.

استعمل المعرف k معادلات الحركة التـــي حلت تكاملياً باستعمال طريقة رونج كوتا (Runge-Kutta) للحساب الرقمي من الدرجة الرابعة. وقد تطلّب إنشاء عنصر التحكم، 30000 عملية تكرار فقط عند استعمال المعرف k.

## 3.5.8 تطبيقات التشخيص Diagnosis Applications

### 1.3.5.8 كشف عطل الممانعة العالية في أنظمة الطاقة الكهربائية

إن أخطاء الممانعة العالية هي أخطاء تيار منخفض تحدث في أنظمة الطاقة الكهربائية. وهذه الأخطاء صعبة الكشف لأنها أخطاء تيار منخفض لا تقدح عادة قواطع التيار أو المنصهرات. ويمكن أن تستمر هذه الأخطاء لبعض الوقت دون أن تكتشف، ومن ثَمّ فقد

تؤدي إلى تهديد السلامة العامة. علاوة على ذلك يمكن أن تؤدي إلى ضياع هام في الطاقة الكهربائية.

لهذا تحاول شركات الطاقة الكهربائية منذ سنين عديدة الكشف عن هذه الأخطاء بالأسلوب الزمني. وقد واجهت الطرق الحسابية بعض النجاح، ولكنها افتقرت المقدرة على التكيف مع الوسط المحيط.

درَّب Fernando، الباحث في مجال الشبكات العصبونية عام 1992 [109] شبكة متعددة الطبقات أمامية التغذية MLFF وشبكة التكرار الجزئية RNN لإنجاز مهام الكشف. واستُعملت 24 مجموعة معطيات اختبار، زُوِّدت من شركة الخدمة الكهربائية بولاية تكساس الأمريكية التـــي احتوت قوساً كهربائياً بشدات مختلفة، وعمليات فتح خط، وفتح/إغلاق مجموعة سعوية.

أولاً حوِّلت المعطيات التشابهية إلى رقمية بتردد أخذ عينات قدره 7680 هرتز، ثم أنجز تحويل فورييه السريع FFTعلى كل إطار معطيات. وأُجري حساب طاقة 128 مركبة ترددية، وجرى حساب طاقة التوافقيات الزوجية والفردية لكل إطار، والتوافقيات فيما بينهما. وحُسِبت أيضاً طاقة تيار التردد العالي المرشح في حقل الزمن، ثم عملت هذه الكميات الأربعة كدخل للشبكات.

وبغية الاختبار، جرى اختيار 28 مقطع معطيات بـــ 1500 عينة من سبعة ملفات معطيات خطأ مرحلي. تضمنت هذه المقاطع عشر معطيات عادية، وثلاث عمليات فتح خط، وثلاث عمليات فتح/إغلاق مجموعة سعوية، و12 خطأ ممانعة عادية.

كانت الشبكة قادرة على الإنجاز بمعدل نجاح بلغ 100% في كشف أخطاء الممانعة العالية. ولا تزال الأبحاث جارية بخطى حثيثة على الشبكات العصبونية لكي تعمل في أوساط وبيئات غير قابلة للتنبؤ.

### 4.5.8 تطبيقات تعرف الأشكال Pattern recognition applications

إن تطبيقات تعرف الأشكال باستعمال الشبكات العصبونية الصنعية ضخمة جداً، وفيما يتعلق بتطبيقات الشبكات التكرارية في هذا المجال سنصف تطبيق رؤية، وتطبيق تمييز إشارة الكلام.

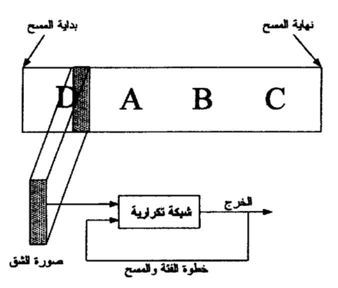
## 1.4.5.8 تعرف سلاسل حرفية جانبية

**Recognitron of lateral character strings**

إن تعرف الأحرف وتمييزها هو أحد الحقول المدروسة بفعالية في تطبيق الشبكات العصبونية الصنعية. وقد أُنجز تقدم كبير في هذا الحقل يتضمن تمييز أحرف الكتابة اليدوية، حيث أُنجزت دقة تمييز من 95% حتـــى 98%.

سنصف هنا تقريباً جديداً لتمييز الأحرف المطبوعة، وذلك بمسح سلسلة جانبية من الحروف. يمكن أن يحدث هذا النوع من القضايا في عملية التصنيع مثلاً، حيث يجب أن تحدد مواصفات الأجزاء بأحرف مطبوعة على جسم تلك الأجزاء. تكون الأحرف عادة مشوهة وفقاً للأسطح غير المنتظمة، والأوساخ، وقلة شروط الإضاءة.

التقريب الذي استعمله Imai عام 1991 [111] هو مسح صورة السلسلة الحرفية من اليسار إلى اليمين باستعمال شق ضيق يمثل دخل شبكة التكرار الديناميكية. نُفِّذ المسح بخطوات، كل خطوة بأربعة عناصر صورة حيث كان حجم صورة الشق 6 × 4 عنصر صورة. عند كل مكان مسح، درِّبت الشبكة لتمييز: هل تنتمي الصورة فعلياً إلى حرف ما؟ أي هل تشكل جزءاً من خطوة تشكيل المقاطع؟ وما هو ذلك الحرف أو ما هي خطوة التمييز؟




**الشكل 20.8**: نظام تعرُّف الأحرف المتتابعة

كان هناك 30 مدخلاً للشبكة (24 + 5 + 1) تتألف من عناصر صورة الشق ومخارج

التغذية العكسية للأحرف المميزة (5 في هذه المجموعة للتجارب)، وخرج إشارة خطوة المسح. أما مخارج الشبكة فقد تألفت من خمس فئات حرفية وإشارة خطوة المسح. وكان عدد العقد المخفية محدداً تجريبياً بـــ 120 عقدة، وقد عُدِّلت الأوزان على وصلات الطبقة المخفية وعقد طبقة الخرج، أما الأوزان على وصلات التغذية العكسية فكانت ثابتة. إن نظام التمييز هذا موضح في الشكل (20.8).

خلال التدريب كان الشق متوضعاً في أقصى اليسار، ومتنقلاً من اليسار إلى اليمين عند كل حرف في السلسلة ليكون متعلماً ومقدماً إلى الشبكة. قورنت الفئة الصحيحة وإشارة الخطوة مع مخارج الشبكة، ثم استعمل الخطأ لخوارزمية الانتشار الخلفي لتعديل الأوزان. ووُضعت إشارات الفئة والخطوة بقيم صفرية عند بداية كل مسح، وبعد إتمام مسح كل حرف.

أُجري التعليم لـــ 4500 عملية تكرار، وبعد تدريب الشبكة فُحِص النظام على عدد من السلاسل الحرفية، فكان مقدار التمييز هو 100% في حالات الاختبار هذه. ثم نُفِّذ اختبار آخر على سلاسل حرفية متصلة يعتريها ضجيج إضافي متضمن في الصورة.

استُعمل لمجموعة الاختبار هذه 52 صورة بـــ 203 حرف. وقد أنجزت الشبكة مقدار تمييز 93% على معطيات الاختبار. ولم يكن هناك أحرف مميزة كفئة خطأ، وإنما كانت الأخطاء هي أخطاء عدم الاستجابة فقط.

تبرهن النتائج المقررة آنفاً مقدرة شبكة التكرار على التقاط المعلومات الزمنية جيداً.

### 2.4.5.8 التوثّق من المتكلم بالاعتماد على النص

**Text-dependent speaker verification**

استُخدمت الشبكات العصبونية الصنعية في التوثّق من شخص المتكلم. يعتبر هذا التطبيق هاماً جداً، حيث يتضمن الوصول إلى تسهيلات الأمان أو المعلومات ورخص المصارف والقروض.

استعمل لمعالجة القضية نموذجان لتمييز المتكلم: نموذج تعريف المتكلم، ونموذج التوثّق من المتكلم. إن تعريف شخصية المتكلم هي عملية تعرُّف شخصية متكلم من بين مجموعة أشخاص متكلمين معروفي طريقة التعبير الكلامية، أما التوثق من المتكلم فهي عملية التوثّق

من شخصية مُدَّعاة لشخص غير معروف.

لقد نفِّذت طرائق عديدة لتعريف الشخصية وللتحقق منها باستعمال التكميم الشعاعي (سيدرس لاحقاً)، ونماذج Markov المخفية HMM (Hidden MarkovModels)، والشبكات المتعددة الطبقات الأمامية التغذية MLFF. نَفَّذ Wang عام 1993 [112] عدة تجارب على بنـــى الشبكات التكرارية باستعمال قاعدة معطيات كلامية مؤلفة من 480 طريقة تعبير.

# آلات بولتزمان ومحاكاة التلدين
# Boltzmann Machines and Simulated Annealing

رأينا في الفصل السابق كيفية اختلاف الشبكات ذات وصلات التغذية العكسية عن الشبكات ذات وصلات التغذية الأمامية فقط. إن ديناميكيات الشبكات ذات التغذية العكسية، كشبكات التكرار الديناميكية، أكثر تعقيداً من سلوك الشبكات الساكنة، حيث توصف ديناميكية هذه الشبكات بمجموعات من المعادلات التفاضلية غير الخطية. ولكن هذه الشبكات أقوى في مجال أنظمة الحساب لأنها قادرة على نمذجة العمليات الزمنية المكانية، بالإضافة إلى أن سلوكيات النظام تكون مستقلة زمانياً أو مكانياً.

سنتابع في هذا الفصل رحلتنا مع الشبكات التكرارية ولكن بآفاق جديدة. فالشبكات المدروسة في هذا الفصل هي شبكات إحصائية، والحالات التي يمكن افتراضها بواسطة الشبكة توصف بتوزيع احتمالي. دُرِس هذا الصنف من الشبكات للمرة الأولى عام 1980، وقَدَّمت هذه الشبكات حلولاً ناجحة لأنواع متعددة من المسائل بما في ذلك الأمثلية التركيبية، والترميز، ومسائل تحويل النصوص إلى كلام، وتخزين الصور، والاستدعاء.

## 1.9 تمهيد

إن آلة بولتزمان هي نوع آخر من الشبكات التكرارية الهامة. درسها Hinton & Sejnowski بين عام 1983[113] و1986[28] وAckley عام 1985[114] وآخرون سنذكرهم لاحقاً. وخلافاً للشبكات المدروسة في المقاطع السابقة فإن آلة بولتزمان هي شبكة إحصائية، والحالات التي تفرض بالشبكة موصوفة بواسطة توزيع بولتزمان؛ وهو عبارة عن شكل أسي لتوزيع احتمالي يُستعمل لنمذجة حالات النظام الفيزيائية عند

التوازن الحراري.

وعلى غرار شبكة هوبفيلد، يوجد لآلة بولتزمان مصفوفة أوزان **W** متناظرة، ولكن من غير المؤكد تقاربها إلى حالة مستقرة، خلافاً لشبكة هوبفيلد.

على أية حال، يمكن أن يكون لآلة بولتزمان وحدات مخفية، وهي الوحدات التـــي لا تتصل مع الوسط الخارجي عند طرفي الدخل والخرج للشبكة. وكذلك هناك مشكلة تعيين الاعتماد التـــي عانت منها الشبكات المتعددة الطبقات الأمامية التغذية في تعيين قيم أوزان الوحدات المخفية عند استعمال التدريب بمعلم.

بالطبع يمكن استعمال خوارزمية الانتشار الخلفي في تعديل أوزان الطبقة المخفية، ولكننا عرفنا في الفصل السادس أن هذه الخوارزمية يمكن أن تنتهي إلى الإخفاق، حيث يمكن أن تقع عملية التعليم في مشكلة الأصغر المحلي وهي في طريق سعيها الحثيث للوصول إلى الأصغر الكلي، وكذلك قد لا تتقارب الخوارزمية نهائياً.

للتغلب على هذه المصاعب، اعتمد الباحثون على بعض المبادئ من فيزياء المادة المكثفة وتطبيقاتها في دراسة الشبكات العصبونية الصنعية فأدى ذلك إلى إيجاد شبكات آلة بولتزمان. طبقت هذه الطريقة، المعروفة بمحاكاة التلدين (simulated annealing)، على الشبكة خلال العمل والتعليم. تسمح هذه الطريقة للشبكة بالانفلات والهروب من مشكلة الأصغر المحلي والتقارب إلى حالة التوازن الكلي.

يمكن أن تعمل آلة بولتزمان في أحد الأطوار الثلاثة المختلفة التالية:

1. يمكن أن تُستعمل كذاكرة مترافقة، وفي هذه الحالة تُستعمل مجموعةٌ مفردة من الوحدات لكل من الدخل والخرج على غرار شبكة هوبفيلد.
2. يمكن أن تُستعمل في مسائل تطبيق الترافق المغاير العامة، على غرار شبكات التغذية العكسية المتعددة الطبقات الأمامية التغذية. عندما تستعمل في مثل هذه المسائل فإن الشبكة تَستعمل وحدات دخل ووحدات خرج منفصلة تدعى الوحدات المرئية.
3. حل مسائل الاستمثال. عند استخدام هذه الشبكات في حل مسائل الاستمثال يكون التعليم غير المتكيف ضرورياً، حيث تعين قيم الأوزان استنتاجياً كجزء من الحل لتابع الكلفة المصاحب للمسألة المعالجة.

## 2.9 خصائص آلة بولتزمان

تعتبر آلة بولتزمان تطويراً لشبكة هوبفيلد، حيث يمكن أن تملك وحدات مخفية بالإضافة إلى وحدات الدخل والخرج. تعمل الوحدات المخفية ككواشف للسمة أو للمَعْلَم الإحصائي لزيادة القدرة الحسابية والتمثيلية لآلة بولتزمان مقارنة مع شبكة هوبفيلد.

البنية الأساسية لهذه الشبكة موضحة في الشكل (1.9)، حيث الوحدات المخفية هي أي وحدات بوصلات داخلية فقط، والوحدات غير المخفية أو المرئية هي وحدات الدخل أو الخرج أوكلاهما معاً.

بوجه عام، لا تكون الشبكة متصلة داخلياً اتصالاً تاماً، لذا قد يكون بعض الأوزان بقيمة الصفر؛ $w_{ij}=0$ ، على أية حال كل الوصلات غير الصفرية تكون متناظرة $w_{ij}=w_{ji}$ ، أي وزن الوصلة من الوحدة $i$ إلى الوحدة $j$ يساوي وزن الوحدة من $j$ إلى $i$، لذلك تبدو هذه الوصلة وكأنها ثنائية الاتجاه بين الوحدة $i$ والوحدة $j$. وكذلك ليس للوحدات تغذية عكسية ذاتية، لذا سيكون $w_{ii}=0$ . ويمكن أن يكون للوحدات دخل انحياز ثابت أو بوجه مكافئ حد عتبة في الدخل التركيبـــي net لكل وحدة.

يعطى الدخل التركيبـــي للوحدة $j$ بـــ :

$$net_j=\sum_{i\in S_j} w_{ij}y_i-\theta \tag{1.9}$$

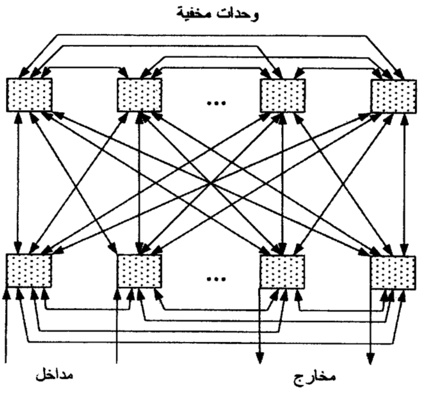
حيث $S_j$ مجموعة الوحدات المتصلة مع الوحدة $j$، و $w_{ij}$ الأوزان التـــي تصل الوحدة $i$ بالوحدة $j$، و $y_i$ توابع تفعيل الخرج الثنائي للوحدات $i$، و$\theta$ حد ثابت العتبة الذي يمكن أن يكون بقيمة الصفر.

تُعيَّن حالة الوحدة $j$، التـــي يمكن أن تكون ثنائية (0, 1) أو ثنائية القطبية (1- ,1+)، احتمالياً كما يلي:

$$y_i=\begin{cases}+1 & p(net_j)\\ -1 & 1-p(net_j)\end{cases} \tag{2.9}$$

حيث يعطى الاحتمال $p(x)$ بواسطة تابع التوزيع من الشكل:

$$p(x)=\frac{1}{1+e^{-x/T}} \tag{3.9}$$




**الشكل 1.9:** شبكة آلة بولتزمان نموذجية

إن T في المعادلة السابقة هو وسيط التحكم بدرجة الحرارة، وسيُعدَّل خلال عمل الشبكة. سنصف فيما بعد دور هذا الوسيط الذي هو اصطلاح فيزيائي مأخوذ من النموذج المستعمل في فيزياء المواد، فهو لا يؤدي المفهوم الحراري في الشبكة العصبونية، وإنما هو وسيط يؤدي دوراً هاماً في تقارب الشبكة. قبل البدء في مناقشة عمل الشبكة سنناقش خوارزمية التعليم بمعلم.

## 3.9 تعليم آلة بولتزمان

كما ذكرنا من قبل، عملية تعليم آلة بولتزمان هي عملية إحصائية. ويجري تعديل الأوزان باستعمال نموذج معدل عن تعليم قاعدة Hebb. عملية التعليم مبنية على النموذج المستعمل في فيزياء المادة المكثفة والمعروف بالتلدين (Annealing). لذا قبل الخوض في خوارزمية تعليم الشبكة سنقوم بإعطاء عرض مختصر عن محاكاة التلدين لتوضيح عملية التعليم وعمل الشبكة.

## 1.3.9 محاكاة التلدين Simulated Annealing

استعير اسم محاكاة التلدين من الفيزياء أُمّ العلوم العلمية، مقارنة مع تلدين المواد الصلبة في الفيزياء. في عملية التلدين توضع المادة الصلبة في ممر حرارة بحيث تزداد درجة الحرارة باستمرار حتـــى تنصهر المادة الصلبة وتتبعثر جزيئاتها فيزيائياً فتصبح متوضعة بترتيب عشوائي. يشار إلى توجيه الجزيئات "بالدوّامات" (Spins). عند هذا المستوى العالي للطاقة، يُبدأ بتبريد ممر الحرارة ببطء وذلك بتخفيض درجة الحرارة T للسماح للجزيئات بالاصطفاف ذاتياً في بنية تصالبية بلورية مرتبة. توافق هذه البنية النهائية حالة طاقة منخفضة مستقرة. بالطبع يجب أن تخفض درجة الحرارة ببطء لتسمح للمادة الصلبة بالوصول إلى التوازن بعد كل انخفاض في درجة الحرارة، وإلا سيحدث اصطفاف غير مرغوب به ينتج عنه عيوب وأخطاء تؤدي إلى حصول " التجمد" إلى الصلب. وهذا يمكن أن يعطي بنية شبه مستقرة عوضاً عن بنية بطاقة منخفضة مستقرة مرغوب بها.

إذا رمزنا لطاقة المادة الصلبة E في الحالة k بـ Ek، فإن التوازن الحراري يحدث في الحالة k باحتمال معين بواسطة توزيع بولتزمان المعرف كما يلي:

$$\Pr(E = E_k) = \frac{1}{Z(T)} \cdot \exp(-E_k / T k_\beta) \qquad (4.9)$$

حيث $k_\beta$ ثابت بولتزمان وقيمته تساوي $1.38 \times 10^{-23}$ جول/ كيلفن، وتابع التجزيء Z(T) هو عامل المعيارية لجعل كتلة الاحتمال الكلي مساوية للواحد. وهكذا:

$$Z(T) = \sum_k \exp(-E_k / T k_\beta) \qquad (5.9)$$

بأخذ المجموع في المعادلة (5.9) عبر كل الحالات الممكنة $2^N$ (بافتراض النموذج الثنائي مع N جزيء).

في محاكاة عملية التبريد، ولـــدت متتالية الحالات من خلال طريقة مونت كارلو (Monte-Carlo) لاختيار الاحتمال. كل حالة k للصلب موافقة لمكان صف ما لكل الجزيئات. البداية من حالة صف عشوائي أولي، ومن ثم يطبق تشويش أو اضطراب صغير بجعل إزاحة لجزيء اختير عشوائياً (في النموذج الثنائي، الاضطراب هو تغير الحالة الثنائية). إذا كان تغير الطاقة الناتج $\Delta E$ تبعاً للإزاحة أصغر من الصفر ( $\Delta E < 0$ الطاقة تتناقص)،

فإن الحالة الجديدة تكون مقبولة.

وإذا كان تغير الطاقة تبعاً للإزاحة أكبر من الصفر ($\Delta E > 0$) فإن الحالة الجديدة تكون مقبولة باحتمال معطى بواسطة:

$$P = \frac{1}{1 + \exp(-\Delta E / Tk_{\beta})} \qquad (6.9)$$

تتكرر عملية الاضطراب حتـــى يصل النظام في آخر المطاف إلى التوازن وتوزيع احتمال الحالة يكون معطى بالعلاقة (4.9).

نلاحظ أنه عندما تكون درجة الحرارة T عالية جداً و$\Delta E > 0$ فإنه يمكن أن يحدث زيادة في الطاقة باحتمال قريب من النصف (1/2). من ناحية أخرى، عندما تكون T قريبة من الصفر فإن الحالة الجديدة تكون محسوبة نهائياً ومعينة.

تسلك الإزاحة سلوكاً مشابهاً للحالة المتقطعة 0/1. وهكذا تسهِّل عملية التلدين نقصان حالة الطاقة، ولكنها تسمح بالزيادة وفقاً للمعادلة (6.9).

عند درجات حرارة عالية، ستكون زيادة الطاقة أكثر احتمالاً للحدوث. وكلما اقتربت درجة الحرارة من الصفر فإن زيادة الطاقة تصبح أقل فأقل احتمالاً. تعتبر المعادلة (6.9) قاعدة القرار المحلي لأن التغير في الحالة يعتمد فقط على الجزيء المضطرب (الوحدة). من المؤكد أنه عندما يكون التوازن الحراري محققاً، فإن حالة الشبكة تخضع لتوزيع بولتزمان المعطى بواسطة العلاقة الكسرية الاحتمالية التالية:

$$\frac{P(\alpha)}{P(\beta)} = \exp[-(E(\alpha) - E(\beta))/T] \qquad (7.9)$$

حيث $\alpha$ و$\beta$ حالات معطاة، و$E(\alpha)$ هي طاقة الحالة $\alpha$، و$E(\beta)$ هي طاقة الحالة $\beta$. بأخذ اللوغاريتم الطبيعي لكلا طرفي هذه العلاقة يظهر أن الفرق في مستوى الطاقة بين الحالتين يكون متناسباً مع الفرق في لوغاريتم احتمال الحالات:

$$\log_e[P(\alpha) - P(\beta)] = -[E(\alpha) - E(\beta)]/T \qquad (8.9)$$

تنفذ محاكاة مشابهة لعملية التلدين الموصوفة آنفاً أيضاً في آلة بولتزمان. ولما كانت مصفوفة الوزن متناظرة، فإنه يمكن تعريف تابع الطاقة للشبكة كما في حالة شبكة هوبفيلد. ففي حالات ثنائية، تعطى طاقة الشبكة E كما عرفها Hinton وSejnowski عام 1985

[115] بالعلاقة التالية:

$$E = -\sum_{ij} w_{ij} y_i y_j \qquad (9.9)$$

(بافتراض قيمة العتبة $\theta = 0$، وعدم وجود انحياز، أو وصلات تغذية ذاتية). هذا التابع له قيمة صغرى عندما تكون حالة الشبكة مستقرة؛ إذ تؤدي حالة الشبكة دور الحالة الصلبة للنظام الفيزيائي، وتابع الطاقة استبدل بتابع الكلفة أو تابع الموضوعية.

أصبح وسيط درجة الحرارة وسيط التحكم الذي يخفّض خطوةً بخطوةً لتقليلٍ بطيء في احتمال الحالة بالتوافق مع جدولة التلدين (Annealing schedule).

بعد فهم عملية التلدين واستحضارها في الذهن سنكون قادرين على البحث في عملية تعليم آلة بولتزمان.

الشكل الأول لتعليم آلات بولتزمان هو التعليم بمعلم حيث تتألف مجموعة التدريب من نماذج الدخل $\mathbf{x}^p$ ونماذج الخرج المنشود الموافقة $\mathbf{t}^p$ وذلك من أجل أزواج النماذج $\{\mathbf{x}^p, \mathbf{t}^p\}$ $p = 1, 2, ...p$ كما ذكر من قبل، يمكن أن تدرب الشبكة لتنفيذ تطبيقات الترافق الذاتي أو المغاير.

في حالة الترافق الذاتي، سيكون النموذج الهدف هو تماماً نموذج الدخل، وتستعمل الشبكة لاستعادة النماذج المخزنة عندما يقدم إلى الشبكة نموذج الدخل الناقص أو الكامل جزئياً (الضجيجي).

أما عندما تدرب الشبكة لإنجاز تطبيقات الترافق المغاير، فإن نماذج الهدف المنشود والدخل تكون على العموم مختلفة. من أجل المناقشة الحالية سنفترض أن للشبكة N وحدة، منها n وحدة دخل، وm وحدة خرج، وh وحدة مخفية $(N = n + m + h)$. عندما نرغب بالإشارة إلى أي وحدة مرئية دون النظر إلى وظيفتها(كدخل أو كخرج) سنستعمل $z_i$ للدلالة على الوحدة المرئية رقم i، حيث $i = 1, 2, .., n + m$.

وبغية الملاءمة الرياضية، سنفترض أن قيم تفعيل الوحدات والدخل ثنائية القطبية $(-1, +1)$.

سنصف أولاً عملية التدريب وصفاً متسلسلاً في تطبيقات الترافق المغاير. وسنجد أن عملية التعليم مملة نوعاً ما لأنها تنفّذ العديد من الخطوات العملية قبل أن تستطيع تحديث

الأوزان. في كل نموذج تدريب، تطبق محاكاة التلدين حتـــى الوصول إلى الحالة المستقرة. ويتحقق هذا في الوقت الذي تكون فيه وحدات الدخل والخرج ملزمة (مثبتة) بقيم نموذج الدخل ونموذج الهدف المنشود. يلي كلَّ إجراءِ عمليةِ تلدينٍ تخزينُ حالة الشبكة عندما يتم تحقق التوازن.

يجمع الإحصاء على حالات الشبكة المخزنة لاستعماله في تقديرات احتمالات التوازن عندما تكون كل الوحدات المرئية ملزمة. بعد ذلك تتكرر العملية ككل فقط عندما تكون وحدات الدخل ملزمة (clamped) أما وحدات الخرج فيسمح لها بتغيير حالتها أو التنفيذ بأسلوب حر (free). تقدر احتمالات حالة التوازن المنفذة بأسلوب حر أيضاً من الحالات المخزنة. ومن ثم تعدَّل الأوزان باستعمال تدرج الهبوط المبنـــي على قياس نظرية المعلومات. تتكرر العملية بعدئذ حتـــى تتقارب الأوزان عبر كل مجموعة التدريب. سيعطي الإجراء عموماً فيما يلي:

### 2.3.9 خوارزمية تعليم بولتزمان

1. وضع كل الأوزان بقيم عشوائية صغيرة، ووضع تفعيلات الوحدات المخفية بقيم عشوائية ثنائية القطبية أولية، ووضع وسيط التحكم بدرجة الحرارة بقيمة أولية عالية $T_0$.
2. لكل زوج من نماذج التدريب دخل،خرج $(\mathbf{x}^p,\mathbf{t}^p)$، نقدم النماذج المختارة إلى الوحدات المرئية للشبكة. النموذج $\mathbf{x}^p$ ملزم بوحدات الدخل والنموذج $\mathbf{t}^p$ ملزم بوحدات الخرج، أما الوحدات المخفية فيسمح لها بتغيير حالتها.
3. يجري اختيار وحدة مخفية كيفياً، ولتكن الوحدة k، وتُغيَّر حالتها من $y_k$ إلى $y'_k$، لاحظ أن $y'_k=-y_k$. ستعطي هذه الحالة المتغيرة تغيراً في طاقة الشبكة $\Delta E_k$، يمكن من المعادلة (9.9) وتناظر الأوزان كتابة :

$$\Delta E_k = E'_k - E_k = -\left[\sum_i w_{ik} y_i (y'_k - y_k) - w_{kk} y_k y'_k\right] \qquad (10.9)$$

حيث يتغير المجموع على $i$ عبر كل الوحدات. وبسبب أننا لم نعتمد وصلات تغذية عكسية ذاتية فإن العلاقة (10.9) تصبح:

$$\Delta E_k = E'_k - E_k = \pm 2\sum_i w_{ik}\ y_i \qquad (11.9)$$

(تذكر أن $y_k = -y'_k$). الإشارة + توافق تغير الوحدة من −1 إلى +1 والإشارة − عندما يكون العكس صحيحاً. لاحظ أنه ما عدا ثابت المضاعفة في العلاقة (11.9)، فهي تماماً الدخل الكلي التركيبـــي المثقل للوحدة k.

إذا كان $\Delta E_k < 0$ ضع الوحدة k بقيمة الواحد بقطع النظر عن حالة تفعيلها السابق (الطاقة تتناقص) ، وإذا كان $\Delta E_k > 0$ ضع تفعيل الوحدة k بقيمة الواحد باحتمال:

$$P_k = \frac{1}{1+\exp(-\Delta E_k / T)} \qquad (12.9)$$

يمكن أن ينجز هذا باستجرار عينة عشوائية من توزيع منتظم وجعل التغير إذا كان $P_k > U$ وإلا تعود الوحدة k إلى حالتها الأصلية.

4. كرر الخطوة 3 لـــ m اختيار وحدة بحيث تغير كل الوحدات المخفية حالتها مرة واحدة وسطياً (السماح للشبكة بالاسترخاء وفقاً للمعادلة (9 - 4). زيادة عدد التكرار t = t + 1.

5. تقليل درجة الحرارة وفقاً لجدولة تلدين ما. كجدولة معدل تخامد أسي بسيط يعطى بـــ:

$$T_{t+1} = \beta\ T_t \qquad (13.9)$$

حيث $T_t$ درجة الحرارة عند الخطوة t و $0 < \beta < 1$ هو معدل التبريد الثابت.

6. كرر الخطوات من 3 إلى 5 حتـــى يتم الوصول إلى درجة الحرارة النهائية $T_{final}$. يكون النظام متوازناً عند هذه النقطة التـــي تكون عندها قيمة E صغرى.

7. خزن حالات كل الوحدات المخفية في نموذج التدريب الملزم p في شعاع $r_c^p$ لاستعمال لاحق في تقدير احتمالات حالة الوحدة.

8. عند إتمام الخطوات السابقة لكل نماذج التدريب، احسب التقديرات $r_{ij}^c$ للارتباطات $\rho_{ij}^c$ لكل أزواج الوحدات التـــي لها نفس الحالات باستعمال الإحصائيات المخزنة $r_c^p$ حيث p =1, 2, .., P، وسيكون:

$$r_{ij}^c = \frac{1}{P}\sum_{p=1}^{P}\varphi\left(z_i^p, h_{ij}^p\right) \quad i,j(i \neq j) = 1,2,..,N \qquad (14.9)$$

وحيث $\varphi(x,y) = 1$ في حالة $x = y$ و $\varphi(x,y) = 0$ ماعدا ذلك .

9. كرر الخطوات من 2 إلى 8 ثانيةً، لكن هذه المرة بدون إلزام وحدات الخرج بقيمة الهدف المنشود ، حيث سيسمح لوحدات الخرج التنفيذ بأسلوب حر. باستعمال الأشعة المخزنة $\mathbf{r}_c^p$ لتقدير ارتباطات التنفيذ الحر، احسب التقديرات $r_{ij}^c$ للارتباطات $\rho_{ij}^c$ باستعمال:

$$r_{ij}^f = \frac{1}{P}\sum_{p=1}^{P}\varphi\left(z_i^p, h_{fj}^p\right) \quad i, j(i \neq j) = 1,2,.., N \qquad (15.9)$$

10. حدِّث أوزان الشبكة $w_{ij}$ وفقاً للقاعدة:

$$\Delta w_{ij} = \alpha\left(r_{ij}^c - r_{ij}^f\right) \quad i, j(i \neq j) = 1,2,.., N \qquad (16.9)$$

حيث $\alpha$ ثابت التعليم، و $r_{ij}^c$ و $r_{ij}^f$ هما الارتباطات المقدرة (احتمالات الحدوث المتبادل) لكلا الوحدتين $i$ و$j$ في شروط الإلزام وغير الإلزام على الترتيب.

11. كرر كل العملية لكل $i$ و$j$ و$p$ حتـــى يكون تغير الأوزان المعطى بالمعادلة (16.9) مساوياً للصفر أو صغيراً بقدر كافٍ.

**مناقشة :**

هناك عدة نقاط تستحق الملاحظة بالنظر إلى الخوارزمية المذكورة آنفاً. في الخطوة 2، يمكن أن يكون هناك فعلياً عدة نماذج خرج منشودة مرتبطة مع نموذج دخل واحد. مثلاً، يحصل ذلك عندما توجد علاقة عائمة (غامضة/مبهمة) أو غير معينة بين نماذج الدخل والخرج كما في حالة النظام الخبير التشخيصي. في هذه الحالة، ستكون نماذج الخرج المنشودة مقدمة إفرادياً مع نموذج الدخل عدداً من المرات الموافقة إلى العلاقات غير المعينة نسبياً لنماذج الخرج.

لا تؤكد الخطوة 3 أن الطاقة ستزداد فعلياً بين الفينة والأخرى، وبذلك يسمح للنظام الهروب من الأصغر المحلي عندما يتحرك أسفل مشهد الطاقة. وهذا يضمن التقارب بوجه محتمل إلى الأصغر الكلي، على الأقل، بأسلوب متدرج. لقد أثبت **Gemen** عام 1984[116] أنه إذا حققت جدولة التبريد العلاقة:

$$T_t \geq \frac{T_o}{1+\log t} \qquad (17.9)$$

لكل t، مع $T_o$ كبير كفاية، عندئذ سيكون التقارب مضموناً تدريجياً. بالطبع يصعب

إنجاز ذلك عملياً. لكن عادة، تَقترح عملية محاكاة التعليم أن التقارب سيحدث عند مستويات الأصغر المحلي مقبولة.

هناك عاملان هامان في عملية التلدين هما جدولة التبريد وعملية الاسترخاء. ستكون درجة الحـــرارة الأوليـــة عاليـــة كفاية بحيث تكون أكثر حالات الشبكة متساوية الاحتمال ($P_k \cong 1/2$ في المعادلة (12.9))، بعدئذ يجب أن تتناقص ببطء . يسمح هذا النوع من الجدولة للشبكة بإنجاز استكشافات كلية كثيرة لمشهد الطاقة الأصلي، ويسمح لها بالهروب من مشكلة الأصغر المحلي. كلما تقدمت عملية التلدين، يزداد المعدل الذي عنده تخفض درجة الحرارة، بعدئذ سينقص احتمال الانتقالات معطياً حالات طاقة منخفضة. عندما تقترب T من الصفر، يقترب احتمال قبول حالة الطاقة من الواحد. تجد عملية الاسترخاء بصفة تكرارية حالة التوازن عند درجة حرارة جديدة باستعمال الحالة النهائية للشبكة ودرجة الحرارة السابقة كنقطة بداية.

بوجه أساسي يجري الوصول إلى التوازن خلال التلدين قبل إنقاص درجة الحرارة وإلا فإن الأصغر الكلي لا يمكن أن يوجد. وقد ثبت أنه عندما يكون عامل التعديل α صغيراً، تكون قاعدة تحديث الأوزان (المعادلة (16.9)) مكافئة لتدرج الهبوط على قياس الأنتروبـــي النسبـــي G لنظرية المعلومات (راجع الأنتروبـــي النسبـــي في الفصل الثالث) بين التوزيعات الملزمة وغير الملزمة (التنفيذ الحر).

إذا كان $P_{ij}^c$ يشير إلى الاحتمال المشترك للوحدات $i$ و$j$ الموجودة بوضع on عندما تكون فقط الوحدات المرئية ملزمة، و $P_{ij}^f$ هو الاحتمال المشترك للوحدات $i$ و$j$ الموجودة بوضع on عندما تكون وحدات الدخل ملزمة وينفذ الخرج بأسلوب حر (وحدات الخرج غير ملزمة)، فإن الأوزان ستعدل بحيث يبقى أحد التوزيعين أقرب ما يكون من الآخر. إحدى الطرائق لتحقيق ذلك هي بتعديل الأوزان لجعل الأنتروبـــي النسبـــي للتوزيعين ذا قيمة صغرى أو إنقاص المسافة بينهما.

نستطيع إنجاز ذلك بتنفيذ تدرج الهبوط على الأنتروبـــي النسبـــي لتعديل الأوزان تعديلاً

متناسباً مع التدرج السالب للأنتروبـــي النسبـــي أي:

$$\Delta w_{ij} = -\alpha\, \partial G / \partial w_{ij} \qquad (18.9)$$

حيث يعرف الأنتروبـــي النسبـــي بين $p^c$ و $P^f$ بما يلي:

$$G\left(P^c \| P^f\right) = \sum_{i,j} P_{ij}^c \log \frac{P_{ij}^c}{P_{ij}^f} \qquad (19.9)$$

وينفذ المجموع عبر كل $i, j = 1,2,...,N$ . يمثل $P_{ij}^c$ احتمالات الحالة المنشودة للشبكة، ويمثل $P_{ij}^f$ احتمالات الحالة الفعلية للشبكة. إن قيمة G موجبة ما لم يكن التوزيعان متماثلين عند النقطة التـــي عندها سيكون G = 0. لذا، يوفر حل المعادلة (18.9) قاعدة تحديث الأوزان التـــي تميل إلى إحضار احتمالات الحالة المنشودة والفعلية معاً. يمكن أن يوجد الحل بأخذ المشتقات الجزئية لـــ G بالنسبة للأوزان والقيام بالتعويضات المناسبة للحصول على قاعدة التحديث (16.9).

عند أخذ المشتقات الجزئية في المعادلة (18.9) يجب ملاحظة أن $P_{ij}^c$ مستقل عن الأوزان لأن وحدات الدخل تكون ملزمة بقيم نموذج الدخل، أما $P_{ij}^f$ فإنه يعتمد على الأوزان. على أية حال، يجب أن يستعمل توزيع بولتزمان (المعادلة (4.9)) لتعيين قيم الأوزان. بعد إيجاد قيم الأوزان لمجموعة نماذج التدريب، تستعمل الشبكة لمهام التطبيق (mapping) غير المعروفة. في هذه الحالة سيكون التذكر (أو عملية التطبيق) أبسط نوعاً ما من خوارزمية التعليم المطولة، وسنعمد الآن لإنشاء هذا التطبيق مباشرة.

### 3.3.9 خوارزمية تطبيق بولتزمان

يمكن أن تستعمل شبكة مدربة معطاة لإنجاز تطبيق (mapping) مرغوب به بتقديم نموذج الدخل إلى وحدات الدخل. يستمر تلدين الشبكة حتـــى تستقر، ثم تُقرأ وحدات الخرج. كما في حالة التعليم، يقاد عمل الشبكة إحصائياً.

تتلخص عملية تطبيق الترافق المغاير كما يلي :

1. ضع كل تفعيلات وحدات الطبقة المخفية والخرج بقيم عشوائية أولية ما (± 1)، وضع

وسيط التحكم بدرجة الحرارة بقيمة عالية $T_0$. ضع (ألزم) نموذج الدخل **x** بوحدات الدخل، واسمح لوحدات الخرج بتغيير حالتها.

2. اختر وحدة خرج أو وحدة مخفية عشوائياً، ولتكن الوحدة *j*، واحسب الدخل التركيبي المثقل **net** لهذه الوحدة :

$$net_j = \sum_{i \in S_j} w_{ij} y_i \qquad (20.9)$$

3. بقطع النظر عن الحالة الجارية للوحدة j، ضعها on (+1) مع احتمال:

$$p(net_j) = \frac{1}{1 + \exp(-net_j / T)} \qquad (21.9)$$

يُنجَز هذا باستجرار العينة العشوائية **U** من التوزيع المنتظم ووضع الوحدة j بحالة on إذا كان $P > U$ ، وإلا أَرْجِع الوحدة j إلى حالتها الأولية.

4. كرر الخطوات 2 و3 حتـــى يَحدث وسطياً اختيار كل وحدات الخرج والوحدات المخفية، وهذا سينظر له كدور عينة واحدة.

5. كرر الخطوة 4 لعدة أدوار حتـــى يتم الوصول إلى التوازن. قلل درجة الحرارة وفقاً للجدولة التالية:

$$T_{t+1} = \beta \, T_t$$

$$T_t = \frac{T_o}{1 + \log t} \qquad (22.9)$$

حيث $0 < \beta < 1$، و $T_0$ درجة الحرارة الأولية.

6. كرر الخطوات 2 و5 حتـــى تستقر الشبكة عند درجة حرارة منخفضة. ومن ثم يؤخذ الخرج المطبق من وحدة الخرج.

التطبيقات المنجزة باستعمال الخوارزمية السابقة يمكن أن تستعمل لمسائل التصنيف حيث كل k غرضاً يصنف إلى c صفاً. فيما يلي سنناقش آلة بولتزمان عندما تستعمل كشبكة ترافق ذاتـــي من أجل تطبيقات إتمام النموذج الناقص الضجيجي.

## 4.9 آلة بولتزمان لإتمام النموذج

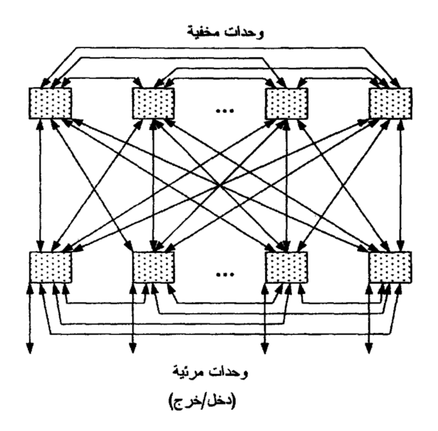
عندما تُستعمل آلة بولتزمان ذاكرة ترافق ذاتي، فإن نماذج الدخل الضجيجي أو الجزئي تكون مقدَّمة إلى الشبكة وتُسترد النماذج الأصلية الكاملة المخزنة. بوجه أساسي، إن العملية هي نفسها كتطبيق ترافق مغاير، ماعدا أنه لا يكون هناك وحدات دخل أو وحدات خرج، بل سيكون هناك فقط الوحدات المرئية التـــي تخدم كمداخل ومخارج. في هذه الحالة ستختلف بنية الشبكة وستكون مشابهة لشبكة هوبفيلد كما هو موضح في الشكل (9-2).

لإجراء عملية استرداد النموذج المخزن من قبل، يلزم وجود نموذج جزئي في الدخل، ثم تُلدَّن الشبكة حتـــى يتم الوصول إلى التوازن. يُسمَح للوحدات بمركبات نموذج الدخل غير المعروفة ووحدات الطبقة المخفية بالتنفيذ الحر خلال عملية التلدين. عند الوصول إلى التلدين يؤخذ الخرج المحسوب من الوحدات المرئية. في تدريب هذا النوع من الشبكات، يتبع إجراء التدريب المشروح آنفاً نفسه لتعليم الترافق المغاير ماعدا نموذج الدخل. النموذج الذي يكون مخزناً، يلزم لكل الوحدات المرئية ويسمح لوحدات الطبقة المخفية بالتنفيذ الحر.

تُتَّبع خطوات جدولة التلدين، وتجمع الإحصاءات على حالات الوحدات غير الملزمة بعد الوصول إلى التوازن، ويستمر تنفيذ الشبكة لأدوار عديدة.

تتكرر نفس العملية بعدئذ بوحدات دخل غير ملزمة (يسمح لها بالتنفيذ الحر)، وتجمع إحصاءات الحدوث المشترك ثانية على حالة الشبكة. تعدل الأوزان وفقاً للمعادلة (9-16) حتـــى الوصول إلى تغير صغير جداً، وتكرر العملية لكل نماذج التدريب. بعد أن تدرَّب الشبكة، يمكنها أن تنجز إتمام نموذج الدخل الضجيجي غير الكامل، وذلك بالعمل كذاكرة عنونة بالمحتوى أو ترافق ذاتي.

من الملاحظ أن أحد العوائق الرئيسية لآلة بولتزمان هو الكمية المفرطة في الحسابات اللازمة لكلٍّ من التدريب والعمل. وهذا ما دفع الكثيرين إلى البحث عن طرق لتسريع العملية ككل. إن آلة كوشي التـــي ستناقش فيما يلي تبدي بعض التحسينات في هذا الاتجاه.




**الشكل 2.9**: بنية آلة بولتزمان لإتمام النموذج

### 1.4.9 آلة كوشي The Cauchy Machine

لشبكات آلة بولتزمان مقدرات تطبيق منافسة للشبكات المتعددة الطبقات الأمامية التغذية (MLFF) وشبكات تابع الأساس الشعاعي. وباستخدام محاكاة التلدين، يمكن تجنب مشكلة الأصغر المحلي التـــي لم تستطيع الشبكات المتعددة الطبقات تجنبها. ولكن لماذا لم تجلب هذه الشبكات نفس الانتباه كالشبكات المتعددة الطبقات الأمامية التغذية؟.

من الواضح أن السبب هو الحسابات اللازمة للوصول إلى التوازن عند كل درجة حرارة، وكذلك جدولة تقليل درجة الحرارة أسيّ. وهذا ما دفع بعض الباحثين مثل Szu & Hartley عام 1987 [117] إلى البحث عن طرق تسريع عملية التلدين. فقد أُنجز بعض النجاح من جهود هذين الباحثَين، حيث استعملا توزيع كوشي بدلاً من توزيع بولتزمان، وقد سمّيا الشبكة الناتجة بآلة كوشي.

يعطى تابع كثافة احتمال كوشي بالمعادلة:

$$f(x,\theta)=\frac{1}{\pi[1+(x-\theta)^2]} \quad (23.9)$$

هذا توزيع متناظر متركز عند θ وله ذيول طويلة جداً و تباين لانهائي (عدم وجود عزوم). يعطي أخذ عينات من هذا التوزيع زيادة في احتمال حجم الخطوة أكثر خلال التلدين. بتعويض توزيع كوشي بدلاً من توزيع بولتزمان، يمكن أن تستعمل جدولة تلدين أسرع.

جدولة تقليل درجة الحرارة هي :

$$T_t = T_o/(1+t) \quad (24.9)$$

لاحظ أن هذه الجدولة **خطية معكوسة** بدلاً من **اللوغاريتمي المعكوس** كما في حالة بولتزمان. وكانت النتيجة النهائية تحسيناً كافياً في عملية التبريد. حتـــى مع زيادة السرعة، فإن زمن التدريب مازال طويلاً جداً (حوالي 100 مرة أكثر) بالمقارنة مع الشبكات المتعددة الطبقات الأمامية التغذية المكافئة المدربة بالانتشار الخلفي.

استُعمل تقريب آخر لتسريع عملية التلدين بُنـــي على مفهوم آخر استعير من الميكانيك يسمى تلدين متوسط الحقل الذي سنفصل معالجته فيما يلي.

### 2.4.9 تلدين متوسط الحقل Mean Field Annealing

كما ذكر من قبل، لم تحظ محاكاة التلدين بشعبية واسعة بسبب كلفة التدريب الحسابية العالية، وهذا هو الدافع وراء البحث الدائم عن طرائق لتسريع عملية التلدين. من بين التقريبات التـــي حققت بعض النجاح هو تلدين متوسط الحقل.

استُعيرت هذه الطريقة من الميكانيك الإحصائي حيث استبدلت العملية الإحصائية للتلدين بقيمة متوسطة واحدة تعيينية. استبُدلت الحالات الاحتمالية المتقطعة في محاكاة التلدين ووضع مكانها قيمها المتوسطة، كما تحسب عادةً بتقريب متوسط الحقل.

تعطي الشبكة بهذه الطريقة التوازن عند درجة حرارة معطاة مرة أو مرتين أسرع منه في حالة محاكاة التلدين. بعدئذ يجب علينا استبدال قاعدة تحديث الأوزان المعطاة بالمعادلة (16-9) ووضع قاعدة القيمة المتوسطة المناسبة التـــي تحَذف عدة حسابات طويلة لازمة في إيجاد احتمالات الحدث المشترك.

المعادلات (14.9) و(15.9) هي تقدير متوسط قيمة احتمالات الحدث المشترك $E(y_i y_j)$ لمعدل الشبكة ككل عبر مجموعة التدريب في حالتَـــي الإلزام والتنفيذ الحر، على الترتيب. نحصل على هذه التقديرات بعد تخزين حالات التوازن للشبكة الملزمة والتنفيذ الحر، وذلك بعد تدخل حسابات عديدة. إذا حصلنا على تقدير القيم المتوسطة بدون جمع إحصاءات الحالة، يمكننا الوصول إلى زيادة في سرعة زمن التعليم.

إذا أخذنا وحدة مفردة في الشبكة، ولتكن الوحدة $j$، فإن حالة هذه الشبكة تحدَّد احتمالياً بالمعادلة (2.9)، لذا فإن معدل أو متوسط قيمة حالة الوحدة سيكون :

$$E(y_j) = (+1)p(net_j) + (-1)[1 - p(net_j)] \\ = 2p(net_j) - 1 \qquad (25.9)$$

حيث

$$p(net_j) = \frac{1}{1 + \exp(-2net_j / T)} \qquad (26.9)$$

ومن المعادلة (25.9)، نجد

$$E(y_i) = \frac{1 - \exp(-2net_j / T)}{1 + \exp(-2net_j / T)} \\ = \tanh(net_j / T) \qquad (27.9)$$

حيث tanh $(x)$ هو تابع الظل القطعي لــ $x$. أيضاً من (9-20) نضع $m_j = E(y_j)$، وسيكون لدينا:

$$E(net_j) = \sum_i w_{ij} E(y_j) \\ = \sum_i w_{ij} m_j \quad , j = 1,2,..,n \qquad (28.9)$$

هذه المتوسطات هي لوحدات معزولة مفردة، ومن ثَم تكون قيماً صغيرة عندما نحتاج فعلياً إلى متوسطات حدث مشترك $E(y_i y_j)$ عبر كل الوحدات $i \neq j$. نستطيع أخذ تقريب بسيط، واستعمال جداء متوسطات قيم الحالة مع تقدير متوسط جداءات قيم الحالة، أي سنفرض أن :

$$E(y_i y_j) \cong E(y_i) E(y_j) \qquad (29.9)$$

مثل هذا التعويض يكون صحيحاً فقط إذا كانت الوحدات مستقلة. ومع ذلك فهو تعويض مفيد كما أثبت Paterson وAnderson عام 1987[28]. فقد أكدا أن هذا التقريب أكثر رسمية على أساس مشتق تلدين متوسط الحقل.

مازال علينا إيجاد حلول لـــ n معادلة غير خطية (27.9). وهذا ينجز بحل النظام التالي من n معادلة تكرارية (Hertz عام 1991[23]):

$$m_j^{new} = E(y_j) = \tanh\left(\frac{1}{T}\sum_i w_{ij} m_i^{old}\right) \quad , j = 1,2,..,n \qquad (30.9)$$

عندما توجد هذه الحلول، يمكن استعمال قاعدة تحديث أوزان متوسط الحقل لتعديل الأوزان وفقاً لـــ :

$$\Delta w_{ij} = \alpha\left[M_{ij}^c - M_{ij}^f\right] \quad , i, j(i \neq j) = 1,2,.., N \qquad (31.9)$$

حيث $M_{ij}^c = m_i^c m_j^c$ و $M_{ij}^f = m_i^f m_j^f$ هي كميات متوسط حقل الحدث المشترك في الحالات الملزمة والتنفيذ الحر على الترتيب.

يُستعمل هذا النوع من التقريب في الميكانيك الإحصائي عندما يستحيل حساب متوسط القيم لعدد ضخم من الجزيئات المتفاعلة. ويعطى ملخص قاعدة التعليم لنظرية متوسط الحقل بما يلي.

### 3.4.9 خوارزمية تعليم نظرية متوسط الحقل

1. ضع كل الأوزان بأعداد عشوائية صغيرة، وضع تفعيلات الوحدات المخفية بقيم ثنائية القطبية عشوائية أولية. ضع وسيط التحكم في درجة الحرارة بقيمة أولية عالية $T_0$.
2. في كل زوج من نماذج التدريب دخل،خرج $(\mathbf{x^p},\mathbf{t^p})$، قدم النماذج المختارة إلى الوحدات المرئية للشبكة. يلزم النموذج $\mathbf{x^p}$ بوحدة الدخل و يلزم النموذج $\mathbf{t^p}$ بوحدة الخرج. وفي حالة متتالية من درجات الحرارة المتناقصة $\{T_0, T_1, T_2 , .., T_{final}\}$، وفق جدولة ما، حل جملة المعادلات (30.9) تكرارياً للحصول على الحلول $m_j$ لكل الوحدات غير الملزمة. تكون الوحدات الملزمة مخصصة بقيم $m_j = \pm 1$ اعتماداً على كون الوحدة on أو off. نحسب عند $T_{final}$ متوسط حقل الاحتمالات للحدث المشترك $M_{ij}^c$

في حالة $i, j(i \neq j) = 1,2,..,n$ .

3. كرر الخطوة 2 باستثناء وحدات الخرج إذ يسمح لها بالتنفيذ الحر. نحسب عند Tfinal احتمالات الحدث المشترك لمتوسط الحقل $M_{ij}^{f}$ في حالة $i, j(i \neq j) = 1,2,..,n$ .
4. كرر الخطوات 2 و3 حتـــى يتم تقديم كل النماذج إلى الشبكة. عدل الأوزان وفقاً لقاعدة تعليم متوسط الحقل:

$$\Delta w_{ij} = \alpha\left[M_{ij}^{c} - M_{ij}^{f}\right] \quad , i, j(i \neq j) = 1,2,..,N \qquad (32.9)$$

حيث $\alpha$ معدل التعليم و $M_{ij}^{c} = m_i^c m_j^c$ و $M_{ij}^{f} = m_i^f m_j^f$ هي الارتباطات المقدرة (احتمالات الحدث المشترك) لكلا الوحدتين $i$ و $j$ بحالة on في حالات الإلزام وغير الإلزام على الترتيب.

5. كرر العملية ككل لـــ $i$ و $j$ و $p$ حتـــى يكون تغير الأوزان المعطى بالمعادلة (32.9) صغيراً كفاية.

عند الاستعاضة عن قاعدة تعليم محاكاة التلدين بهذه القاعدة، نحصل على سرعة إنجاز أعلى بمرة أو مرتين مع **خسارة قليلة في مستوى الإنجاز**.

استعمل Bilbro عام 1988[41] طرق تلدين متوسط الحقل في حل مسألة التجزيء البيانـــي (graph partitioning) وفي التطبيق العملي للأمثلية التركيبية NP-hard. وأظهرت نتائجه سرعة تصل حتـــى 50 مرة مقارنة مع الإنجاز في التلدين القياسي.

### 4.4.9 نموذج سلسلة ماركوف Markov chain Model

يمكن أن تكون عملية التلدين مميزة رياضياً كسلسلة (Markov chain)؛ فسلسلة ماركوف متتالية من التجارب، حيث توافق نتيجة أي تجربة حالة أو هيئة النظام (الشبكة).

إن السمة المميزة لخاصية ماركوف هي أن الحالة الجديدة تعتمد فقط على الحالة السابقة لها، وليس على كل الحالات التـــي قبل سابقتها. هذه الاعتمادات هي وحيدة الخطوة، ويمكن أن يعبَّر عنها باحتمالات المرور $p_{ij}(n)$ . يعطى الاحتمال المشروط لحركة النظام (مروره) إلى الحالة j على تجربة رقم n بكون النظام في الحالة i على التجربة رقم (n−1) أي:

$$p_{ij}(n) = \Pr\{\mathbf{x}(n) = j \mid \mathbf{x}(n\text{-}1) = i\} \qquad (33.9)$$

حيث $\Pr\{A|\mathbf{x}=k\}$ يشير إلى الاحتمال الشرطي لحدث A معطى، حيث = $k\mathbf{x}$، و $\mathbf{x}(n)$ متحول عشوائي.

هناك نظرية واسعة أقيمت لسلاسل ماركوف تطبق على آلة بولتزمان بمحاكاة التلدين. وكانت هذه النظرية حقاً الأساس لإثبات التقارب إلى طاقة الأصغر الكلي التـــي استعملها Gemen وGemen عام 1984[116]. ساعدت هذه النظرية أيضاً على فهم سلوك عملية التلدين.

## 5.9 حل مسائل الاستمثال Solving Optimization Problems

مسائل الاستمثال موجودة في كل مكان، وهي تحدث بطرائق متكررة في مجالات عديدة مختلفة. مثلاً، تريد الشركات جعل الإنتاجية والأرباح أعظمية، وجعل التكلفة والنفايات أصغرية، وكذلك المخاطرة أصغرية مع نمو أعظمي، وهكذا.

وكانت الشبكات العصبونية التـــي استُعملت بنجاح في هذا الحقل هي الشبكات التكرارية وشبكات هوبفيلد. فكلاهما استُعمل لحل عدد من مسائل الاستمثال مثل مشكلة البائع الجوال، ومشكلة الوزراء على رقعة شطرنج (n-queens)، والتلوين البياني،. ..الخ.

تقع هذه المسائل نموذجياً ضمن مسائل (NP-complete) التركيبية، ففي هذه المسائل ينمو فراغ الحل بمعدلات أسية مع حجم وسطاء المسألة. ومن ثَم فإن حلول الاستمثال لهذه المسائل تتطلب، في الأغلب، أعباءً حسابيةً كبيرة.

سنبدأ بأحد تطبيقات الاستمثال وهو استعمال محاكاة التلدين في إيجاد حل أمثلي لتوزيع المفاتيح في لوح مفاتيح آلة كاتبة.

### 1.5.9 توزيع لوح مفاتيح آلة كاتبة Typewriter keyboard Layout

معظم لوحات مفاتيح الآلة الكاتبة تَستعمل ترتيب مفتاح QWERTY، وهذا الاسم هو أول ستة مفاتيح في السطر العلوي من الأحرف. من المعروف أن هذا الترتيب لمفاتيح اللوح ليس أمثلياً لأسباب عديدة:

1. هذا الترتيب يضع توزيعاً غير متساوٍ لعبء الضرب بالأصابع على لوحة المفاتيح، وهو دائماً لمصلحة اليد اليسرى على اليد اليمنـــى، مع أن معظم البشر يستخدمون اليد اليمنـــى وليس اليسرى .
2. يعطي هذا التوزيع للأصابع الضعيفة عبء طباعة أكبر من الأصابع القوية.
3. يستعمل السطر الوسطي أقل من 3/1 زمن الطباعة، وهذا ناتج عن زيادة زمن انتقال الأصابع.
4. لا يأخذ هذا التصميم بالحسبان حقيقة أن تناوب الأصابع لطباعة زوج من الأحرف أسرع من طباعة نفس الإصبع لهذا الز وج.

نود الإشارة إلى أن ترتيب QWERTY جاء بناءً على دراسات إحصائية للغة الإنكليزية تأخذ بعين الاعتبار الثنائيات، لذا فإن الحل الأمثلي الذي وجد باستخدام الشبكات العصبونية ليس بعيداً عن توزيع QWERTY، كما سنرى فيما بعد.

إن تصميم توزيع لوحة مفاتيح أقرب إلى الاستمثال بتحقيق شروط معينة سيزيل نقاط الضعف المذكورة آنفاً ويعطي معدلات أسرع لمقادير الطباعة. مثلاً، سترتب المفاتيح بأخذ خصائص إحصائية للأحرف الإنكليزية للحصول على أزمنة أصغرية لانتقال الأصابع، وك ذلك ستأخذ بعين الاعتبار ترددات حدوث أزواج الأحرف أو حتـــى ثلاثيات الأحرف (digrams/trigrams) في التوزيع، ويجب أن يوزع عبء الطباعة ليعطي اليد اليمنـــى عملاً أكثر من اليسرى.

ولما كان الهدف المنشود هو زيادة سرعة الطباعة، فإن من المفيد التقاط وتلخيص النقاط والاعتبارات المذكورة آنفاً في قياس إنجاز مفرد لاستعماله في هيئات ترتيب لوحات مفاتيح مختلفة. في تصميم اللوحة هناك أكثر من $26! \cong 4\times10^{26}$ تركيبة لترتيبات المفاتيح إذا اعتبرنا فقط 26 حرفاً أبجدياً سيجري البحث في ترتيبها. إن تقييم كلٍّ منها هو بالطبع غير عملي إذا لم يكن مستحيلاً. ومساحة البحث ستكون ضخمة جداً لإجراء مثل هذا التقييم المتعب لكل توزيع ممكن.

فهذا مثال آخر لمشكلة (NP-complete) (Garey & Johnson عام 1979 [118])، المشكلة التركيبية التـــي تنمو أسيّاً مع عدد المفاتيح. من ناحية أخرى، إن ما اقتُرح فيما سبق

من محاسن، يجعلنا لا نحتاج إلى الإلحاح والتأكيد للحصول على هيئة أو ترتيبة أمثلية، بالطبع **إذا وجد حل جيد كفاية مع حسابات أقل.**

ولكي نكون واثقين بأن هذا الحل الجيد سيوجد، فإننا نحتاج إلى طريقة تستكشف لنا أجزاءً مختلفةً عديدة بدون الوقوع في مشكلة الأصغر المحلي. يمكن أن توفر محاكاة التلدين مثل هذا التقريب.

وضع كل من **Light & Anderson** عام 1993 [119] حل محاكاة تلدين لمسألة لوحة مفاتيح الآلة الكاتبة باستعمال تابع كلفة بسيط. التابع مبنـــي على ترددات نسبية لكل أزواج الأحرف الإنكليزية وزمن العبور بين أزواج مفاتيح اللوحة.

إذا اعتبرنا الأحرف الإنكليزية فقط، فسيكون لدينا:

$$C_2^{26} = \binom{26}{2} = \frac{25 \times 26}{2} = 325$$

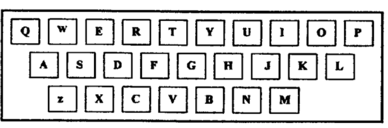
هذه الأزواج يجب أن تؤخذ بعين الاعتبار عند تحقيق توزيعات اللوحة. ستأخذ بعض أزواج الأحرف ترددات صفرية (مثل qx وzq) والبعض الآخر سيأخذ ترددات أعظمية (مثل **ti** و**ch**).

تابع الكلفة المستعمل من قبل **Light & Anderson** عام 1993 [119] هو تقريب لمعدل زمن العبور بين أزواج الأحرف من أجل 325 تركيبة. التابع هو تماماً مجموع الجداءات بين كل ترددات أزوج الأحرف $F_{\alpha\beta}$ وأزمنة العبور الموافقة $T_{pos(\alpha)pos(\beta)}$ بين أزواج الأحرف $\alpha$ و$\beta$ كما يلي:

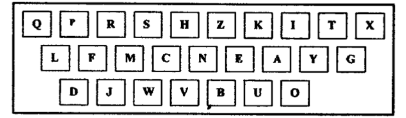
$$\cos t = \sum_{\alpha=a}^{z} \sum_{\beta=a}^{z} F_{\alpha\beta}\, T_{pos(\alpha)pos(\beta)} \qquad (34.9)$$

حيث تحسب الأدلة $\alpha$ و $\beta$ عبر كل الأحرف الأبجدية.

ومع أن تابع الكلفة هذا يتجاهل عوامل سرعة الطباعة، فهو تقريب معقول ومفيد. في إنجاز المحاكاة، تولّدت حلول لتوزيع لوحة مفاتيح مختلفة تماماً عن لوحة **QWERTY**. مثال عن هذه الحلول الأمثلية موضح في الشكل (3.9) حيث قورنت كلفة حل التلدين مع كلفة حل **QWERTY**.




لوحة مفاتيح QWERTY ، بتابع كلفة = 1542




لوحة مفاتيح الحل الأمثل ، بتابع كلفة =1428

**الشكل 3.9**: مقارنة الكلفة وتوزع المفاتيح للوحة QWERTY ولوحة حل محاكاة التلدين الأمثل.

كلفة الحل الأفضل تساوي 1428 بالمقارنة مع كلفة Qwerty التـــي تساوي 1542 (نلاحظ أن الحل الأمثل ليس بعيداً عن حل QWERTY) أي تحقق إنقاص الكلفة بنسبة 7.4% في حلول التلدين. وذلك لتحقيق بعض الخواص المرغوب بها في اللوحة المذكورة آنفاً، بما في ذلك، تخصيص عمل أكبر لليد اليمنى، وتخصيص أزواج حرفية عامة لطباعة الأحرف المتكررة (مثل O, A, T, E وهكذا) بالأصابع القوية، والأحرف الأقل تكرارية خصصت للأصابع الضعيفة.

يمكن لمسألة توزيع لوحة المفاتيح هذه أن تُحلّ بآلة بولتزمان ومحاكاة التلدين إذا خصصنا وسطاء الشبكة (عدد الوحدات والوصلات وقيم الأوزان) بطريقة صحيحة في علاقة تابع الكلفة. هذه المسألة مشابهة لمسألة البائع الجوال التـــي ستشرح بالتفصيل فيما بعد.

ليكن N عدد مفاتيح اللوحة (يقابل عدد المدن في مسألة البائع الجوال) و$D = (d_{ij})$ مصفوفة الجداءات لترددات أزواج المفاتيح وأزمنة الانتقال (وهذا يقابل المسافات بين المدن المطلوب من البائع الجوال المرور عليها). يمكن تعريف المتحول المؤشر $x_{ip}$ كما يلي:

**إذا ضغط المفتاح i في المكان p فإن** $x_{ip} = 0$ **وإلا سيكون** $x_{ip} = 1$ .

بعدئذ نرغب بجعل التابع الموضوعي أو تابع الكلفة التالي أصغر ما أمكن:

$$f(x)=\sum_{i,j,p,q}^{N-1} a_{ijpq}x_{ip}x_{jq} \tag{35.9}$$

والخاضع للشروط المقيدة التالية:

$$\begin{aligned}&\sum_{i=0}^{N-1} x_{ip}=1 \quad , p=0,1,2,..,N-1\\&\sum_{p=0}^{N-1} x_{ip}=1 \quad , i=0,1,2,..,N-1\end{aligned} \tag{36.9}$$

و إذا كان $q=(p+1)\bmod N$ فإن $a_{ijpq}=d_{ij}$ وإلا $a_{ijpq}=0$

لحَلّ آلة بولتزمان عرّف Korst وAarts عام 1989 [120] الوحدات $u_{ip}$ المرافقة لـــ $x_{ip}$ والانحياز، والمسافة والوصلات المُخمَّدة للشبكة. تَستعمل تشكيلة حلّهما شبكةً مؤلفـــة مـــن ($2N^3$-$N^2$) وصلة.

## 2.5.9 مسألة البائع الجوال Traveling Salesman problem

ينبغي على البائع الجوال زيارة **مجموعة معطاة من n مدينة** (وهذه المدن تقابل مجموعة المفاتيح في التطبيق السابق) **مرة واحدة لكل مدينة وفقط مرة**، ومن ثم العودة للمدينة المنطلق (البداية) عند نهاية جولته، بحيث تكون **مسافة الجولة أقصر ما أمكن**.

إن صعوبة إيجاد الحل تزداد سريعاً كلما ازداد عدد المدن، حيث سيكون هناك $(n-1)!$ من الجولات المختلفة، وهذا العدد سيزداد أسياً بزيادة عدد المدن n المطلوب المرور عليها، وهذا مثال آخر لمشكلة (NP-complete)، (Garey & Johnson عام 1979 [xx]).

هناك الكثير من الباحثين الذين ناقشوا هذه المسألة بإسهاب مثل Lawler, Lenstra, Rinnooy Kan , Shmoys عام 1985 [121].

إن الشبكات العصبونية الصنعية تمتاز بميزات قوية مقارنة مع التقنيات الأخرى التقليدية في حل بعض مسائل الاستمثال فهي تستطيع معالجة الحالات التـــي تكون فيها بعض الشروط المقيدة ضعيفة (مرغوب فيها ولكن ليست متطلبة بصفة مطلقة). مثلاً، في مسألة البائع

الجوال، من المستحيل فيزيائياً زيارة مدينتين بنفس اللحظة (شرط مقيد مطلق)، ولكن من المرغوب به زيارة المدينة مرة واحدة فقط (شرط مقيد ضعيف).

سينعكس الفرق بين هذه الشروط المقيدة بجعل عقوبة على الوحدتين اللتين ستكونان on في نفس اللحظة (أي في نفس العمود في مصفوفة المسافات المعطاة فيما بعد) أكبر من العقوبة على الوحدتين اللتين تكونان on في نفس اللحظة (أي في نفس السطر في مصفوفة المسافات). وإذا كانت زيارة بعض المدن أكثر أهمية من زيارة البعض الآخر فيمكن عندها أن تعطى هذه المدن المهمة أوزان وصلة ذاتية أكبر من أوزان الوصلة الذاتية للمدن التـــي هي أقل أهمية .

سنوضح المسألة باستخدام عدة شبكات، وسنرى مقدرتها على إيجاد الحل الأمثل لهذه المسألة بعشرة مدن ($n = 10$)، وبالتالي سيكون أمام البائع $(n-1)! = 362880$ جولة ممكنة فقط!، والتـــي استخدمت من قبل العديد من الباحثين مثل Wilson & Pawley عام 1988 [122] وSzu عام 1988[123].

ستكون إحداثيات المدن كما يلي:

| المدينة | $X_2$ | $X_1$ |
|---|---|---|
| A | 0.4000 | 0.4439 |
| B | 0.2439 | 0.1463 |
| C | 0.1707 | 0.2293 |
| D | 0.2293 | 0.7610 |
| E | 0.5171 | 0.9414 |
| F | 0.8732 | 0.6536 |
| G | 0.6878 | 0.5219 |
| H | 0.8488 | 0.3609 |
| I | 0.6683 | 0.2536 |
| J | 0.6195 | 0.2634 |

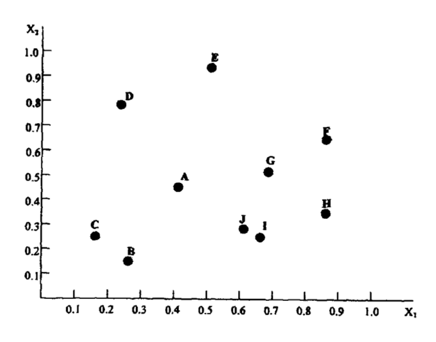
وتعطى المسافات بين المدن على شكل مصفوفة متناظرة لها القيم التالية:

|   | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0.0000 | 0..3361 | 0.3141 | 0.3601 | 0.5111 |
| B | 0.3361 | 0.0000 | 0.1107 | 0.6149 | 0.8407 |
| C | 0.3141 | 0.1107 | 0.0000 | 0.5349 | 0.7919 |
| D | 0.3601 | 0.6149 | 0.5349 | 0.0000 | 0.3397 |
| E | 0.5111 | 0.8407 | 0.7919 | 0.3397 | 0.0000 |
| F | 0.5176 | 0.8083 | 0.8207 | 0.6528 | 0.4579 |
| G | 0.2982 | 0.5815 | 0.5941 | 0.5171 | 0.4529 |
| H | 0.4564 | 0.6418 | 0.6908 | 0.7375 | 0.6686 |
| I | 0.3289 | 0.4378 | 0.4982 | 0.6710 | 0.7042 |
| J | 0.2842 | 0.3934 | 0.4501 | 0.6323 | 0.6857 |

|   | F | G | H | I | J |
|---|---|---|---|---|---|
| A | 0.5176 | 0.2982 | 0.4564 | 0.3289 | 0.2842 |
| B | 0.8083 | 0.5815 | 0.6418 | 0.4378 | 0.3934 |
| C | 0.8207 | 0.594 | 0.6908 | 0.4982 | 0.4501 |
| D | 0.6528 | 0.5171 | 0.7375 | 0.6710 | 0.6323 |
| E | 0.4579 | 0.4529 | 0.6686 | 0.7042 | 0.6857 |
| F | 0.0000 | 0.2274 | 0.2937 | 0.4494 | 0.4654 |
| G | 0.2274 | 0.0000 | 0.2277 | 0.2690 | 0.2674 |
| H | 0.2937 | 0.2277 | 0.0000 | 0.2100 | 0.2492 |
| I | 0.4494 | 0.2690 | 0.2100 | 0.0000 | 0.0498 |
| J | 0.4654 | 0.2674 | 0.2492 | 0.0498 | 0.0000 |

إن مكان المدن وإحداثياتها مبينة في الشكل (4.9) التالي:

فيما يخص الشبكة العصبونية لمسألة البائع الجوال ومن أجل n مدينة سنستعمل $n^2$ وحدة مرتبة $U_{city,position}$ في شكل مصفوفة مربعة ببعد 10 × 10 كما هو موضح في الشكل (5-9). كل وحدة تمثل فرضية، ستكون هذه الوحدة on إذا كانت الفرضية صحيحة، وستكون off إذا كانت الفرضية خاطئة. ستكون الأوزان معينة لتمثيل الشروط المقيدة للمسألة وتابع الكلفة لتحقيق الاستمثال.




**الشكل 4.9**: المدن العشر لمسألة البائع الجوال

سيوافق حل المسألة تابع طاقة أصغري لتابع إجماع أعظمي للشبكة. وسيعدل مستوى الفعالية لكل وحدة حتـــى تجد الشبكة القيمة الصغرى أو العظمى المرغوب بها. ستكون حالات الوحدات بقيم ثنائية مع انتقالات حالة (من حالة إلى أخرى) احتمالية، وستمثل هيئة الشبكة بشعاع حالات الوحدات. الوصف المقدم هنا مبنـــي على تابع الأعظمية أو تابع الإجماع (Aarts & Korst عام 1989[120]) المعطى بالعلاقة التالية:

$$C = \sum_{i} \left[ \sum_{j \le i} w_{ij} x_i x_j \right] \tag{37.9}$$

ينفذ المجموع عبـــر كل وحدات الشبكة، حيـــث $x_i$ حالة الوحدة $X_i$ وستكون إما +1 (on) وإما صفراً (off)، و $w_{ij}$ الأوزان المثبتة التـــي تعبر عن درجة الرغبة بأن تكون كلا الوحدتين $X_i$ و $X_j$ بحالة on.

ستحاول الشبكة إيجاد هذا التابع الأعظمي (أو على الأقل الأعظمي المحلي) بترك كل وحدة تحاول تغيير حالتها من on إلى off أو بالعكس، حيث يمكن أن تنفذ هذه المحاولات إما تسلسلياً (وحدة واحدة كل لحظة) وهو الذي سيعتمد هنا، أو تفرعياً (عدة وحدات في اللحظة الواحدة).

تعتبر الجولة صحيحة تماماً عندما تكون وحدة واحدة on فقط في كل سطر وفي كل

عمود. يعنـــي وجود وحدتين on في السطر أن المدينة الموافقة قد زارها البائع مرتين، ويعنـــي وجود وحدتين on في العمود أن البائع كان في المدينتين بنفس اللحظة.

ستكون الوحدات في كل سطر متصلة داخلياً اتصالاً كاملاً، وبالمثل فإن الوحدات في كل عمود ستكون متصلة اتصالاً كاملاً. تُعدَّل الأوزان بحيث أن الوحدات ضمن نفس السطر (أو نفس العمود) لا تكون on عند نفس الزمن. بالإضافة إلى ذلك، هناك وصلات بين الوحدات في الأعمدة المتجاورة وبين الوحدات في أول وآخر عمود، وفقاً للمسافات بين المدن.

في مناقشتنا لمسألة البائع الجوال سنستعمل آلة بولتزمان وفقاً لأبحاث Aarts وKorst عام 1989 [120] وسنعطي الآن ملخصاً للمصطلحات المستخدمة:

n : عدد المدن في الجولة (وسيكون هناك $n^2$ وحدة في الشبكة)

i : دليل المدينة، حيث $1 \le i \le n$

j : دليل المكان في الجولة ، mod n أي: $j = n+1 \rightarrow j = 1, j = 0 \rightarrow j = n$

$U_{ij}$ : وحدة تمثل أن المدينة رقم i قد جرت زيارتها عند الخطوة رقم j من الجولة

$u_{ij}$: تفعيل الوحدة $U_{ij}$ : $u_{ij} = 1$ إذا كانت الشروط الأولية صحيحة و$u_{ij} = 0$ إذا كانت خاطئة

$d_{ik}$: المسافة بين المدينة i والمدينة k، $k \neq i$

d: المسافة العظمى بين أي مدينتين

من المناسب ترتيب الوحدات في الشبكة العصبونية كشبكية مخططـــة كما فـــي الشكل (5.9). تمثل أسطر الشبكية المدن الواجب زيارتها، وتمثل الأعمدة مكان المدينة في الجولة. كما ذكر من قبل، يجب أن تكون الوحدات ضمن كل سطر وعمود متصلة اتصالاً كاملاً، وستكون الأوزان على كل الوصلات بقيمة p ($p > 0$). بالإضافة إلى ذلك فإن كل وحدة لها وصلة ذاتية بوزن $b > 0$ كما هـــو موضح فـــي الشكل (6.9)، حيث رمزنا للوحدات بـــ $U_{i,j}$.

سيكون نموذج وصل الشبكة العصبونية كما يلي:

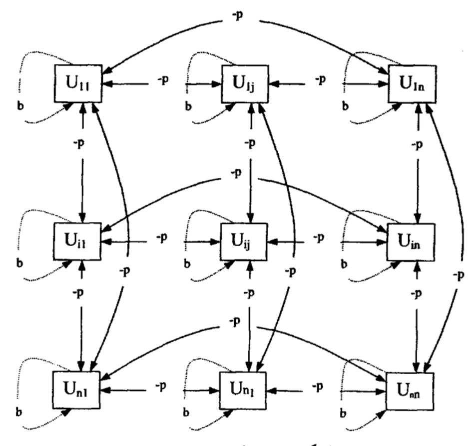
$U_{i,j}$ : لها وصلة ذاتية بوزن b، وهذا يمثل الرغبة بزيارة المدينة $i$ في المرحلة $j$ .

$U_{i,j}$ : موصلة إلى كل الوحدات الأخرى في السطر $i$ بأوزان مفروضة $-p$، هذا يمثل أن نفس المدينة يجب ألاّ تزار مرتين.

| المدينة | المكان | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| A | $U_{A1}$ | $U_{A2}$ | $U_{A3}$ | $U_{A4}$ | $U_{A5}$ |
| B | $U_{B1}$ | $U_{B2}$ | $U_{B3}$ | $U_{B4}$ | $U_{B5}$ |
| C | $U_{C1}$ | $U_{C2}$ | $U_{C3}$ | $U_{C4}$ | $U_{C5}$ |
| D | $U_{D1}$ | $U_{D2}$ | $U_{D3}$ | $U_{D4}$ | $U_{D5}$ |
| E | $U_{E1}$ | $U_{E2}$ | $U_{E3}$ | $U_{E4}$ | $U_{E5}$ |
| F | $U_{F1}$ | $U_{F2}$ | $U_{F3}$ | $U_{F4}$ | $U_{F5}$ |
| G | $U_{G1}$ | $U_{G2}$ | $U_{G3}$ | $U_{G4}$ | $U_{G5}$ |
| H | $U_{H1}$ | $U_{H2}$ | $U_{H3}$ | $U_{H4}$ | $U_{H5}$ |
| I | $U_{I1}$ | $U_{I2}$ | $U_{I3}$ | $U_{I4}$ | $U_{I5}$ |
| J | $U_{J1}$ | $U_{J2}$ | $U_{J3}$ | $U_{J4}$ | $U_{J5}$ |

| المدينة | المكان | | | | |
|---|---|---|---|---|---|
| | 6 | 7 | 8 | 9 | 10 |
| A | $U_{A6}$ | $U_{A7}$ | $U_{A8}$ | $U_{A9}$ | $U_{A10}$ |
| B | $U_{B6}$ | $U_{B7}$ | $U_{B8}$ | $U_{B9}$ | $U_{B10}$ |
| C | $U_{C6}$ | $U_{C7}$ | $U_{C8}$ | $U_{C9}$ | $U_{C10}$ |
| D | $U_{D6}$ | $U_{D7}$ | $U_{D8}$ | $U_{D9}$ | $U_{D10}$ |
| E | $U_{E6}$ | $U_{E7}$ | $U_{E8}$ | $U_{E9}$ | $U_{E10}$ |
| F | $U_{F6}$ | $U_{F7}$ | $U_{F8}$ | $U_{F9}$ | $U_{F10}$ |
| G | $U_{G6}$ | $U_{G7}$ | $U_{G8}$ | $U_{G9}$ | $U_{G10}$ |
| H | $U_{H6}$ | $U_{H7}$ | $U_{H8}$ | $U_{H9}$ | $U_{H10}$ |
| I | $U_{I6}$ | $U_{I7}$ | $U_{I8}$ | $U_{I9}$ | $U_{I10}$ |
| J | $U_{J6}$ | $U_{J7}$ | $U_{J8}$ | $U_{J9}$ | $U_{J10}$ |

**الشكل 5.9**: البنية من أجل عشر مدن في مسألة البائع الجوال




**الشكل 6.9**: شبكة آلة بولتزمان

$U_{i,j}$: موصلة إلى كل الوحدات الأخرى في العمود $j$ بأوزان مفروضة $-p$، هذا يمثل أن مدينتين لا يمكن أن تزارا بنفس اللحظة .

$U_{i,j}$: موصلة إلى $U_{k,j+1}$، من أجل $1 \leq k \leq n$ و $k \neq i$، بوزن $-d_{ik}$، هذا يمثل المسافة المقطوعة للقيام بالعبور من المدينة $i$ عند المرحلة $j$ إلى المدينة $k$ عند المرحلة $j + 1$.

$U_{i,j}$: موصلة إلى $U_{k,j-1}$، من أجل $1 \leq k \leq n$ و $k \neq i$ ، بوزن $-d_{ik}$، هذا يمثل المسافة المقطوعة للقيام بالعبور من المدينة $k$ عند المرحلة $j - 1$ إلى المدينة $i$ عند المرحلة $j$.

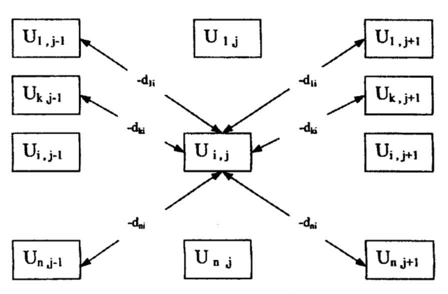
إن الشبكة المرغوب بها ستكوَّن بخطوتين، أولاً، ستكوَّن الشبكة العصبونية لمسألة حدوث الإجماع الأعظمي وذلك متـى تحققت كافة الشروط المقيدة للشبكة، أي عندما تكون وحدة واحدة فقط on في كل سطر وفي كل عمود. ثانياً، سنضيف وصلات مثقلة لتمثيل المسافات بين المدن. لمعالجة مسألة الإجماع الأعظمي، نفترض أن الأوزان الممثلة للمسافات ستكون سالبة. تمثّل آلة بولتزمان بأوزان (p, b) الشروط المقيدة (لكن ليس مسافات) لمسألة

البائع الجوال الموضحة في الشكل (6.9).

إذا كان $p > b$ فإن الشبكة ستعمل كما هو مرغوب به (كما شرح من قبل). ولإتمام عمل شبكة بولتزمان في مسألة البائع الجوال، علينا إضافة الأوزان التـــي تمثل المسافات بين المدن. إن الوحدة النموذجية $U_{i,j}$ موصلة إلى الوحدات $U_{k,j-1}$ و $U_{k,j+1}$ (لكل $k \neq i$) بواسطة الأوزان التـــي تمثل المسافات بين المدينة $i$ والمدينة $k$. أوزان المسافات مبينة في الشكل (7.9) للوحدة النموذجية $U_{ij}$. لاحظ أن الوحدات في العمود الأخير موصلة إلى الوحدات في العمود الأول بوصلات ممثلة للمسافات المناسبة.

الآن سنحاول استكشاف العلاقة بين الوزن المقيد b وأوزان المسافة. لتكن d المسافة العظمى بين أي مدينتين في الجولة. نفترض أنه لن تزار أية مدينة في المكان رقم زمن الجولة، وكذلك لن تزار مدينة مرتين. في هذه الحالة، إذا قلنا: إن مدينة ما ــ ولتكن $i$ ــ لن تزار أبداً، فإن هذا يعنـــي أنه لن تكون أي وحدة on في العمود $j$ أو في السطر $i$.

وحيث إن السماح للوحدة $U_{ij}$ لتعود إلى حالة on سيكون مشجعاً، فإن الأوزان يجب أن توضع بحيث يزداد الإجماع إذا عادت إلى الحالة on. التغير في الإجماع سيكون $b - d_{ik_1} - d_{ik_2}$، حيث $k_1$ يشير إلى المدينة المزارة عند المرحلة $j - 1$ من الجولة و $k_2$ يشير إلى المدينة المزارة عند المرحلة $j + 1$ (والمدينة $i$ مزارة عند المرحلة $j$).




**الشكل 7.9**: شبكة بولتزمان لمسألة البائـــع الجوال، الأوزان الممثلة للمسافات للوحدة Ui,j.

سيكون هذا التغير أكبر أو يساوي b–2d. على أية حال، التساوي سيحدث فقط إذا كانت المدينتان المزارتان في الأماكن $j - 1$ و $j + 1$ بمسافة أعظمية d بعيدة عن المدينة $i$. بوجه عام، يكفي أن يكون تغير الإجماع موجباً، لذا، سنأخذ $b > 2d$. وهكذا نــرى أنه إذا كان $p > b$، فإن تابع الإجماع سيأخذ قيمة أعظمية، في حل ملائم (يحقق الشروط المقيدة)، بالمقارنة مع الحل غير الملائم، وإذا كان $b > 2d$ سيكون الإجماع أعلى، في حل ملائم قصير، منه لجولة أطول.

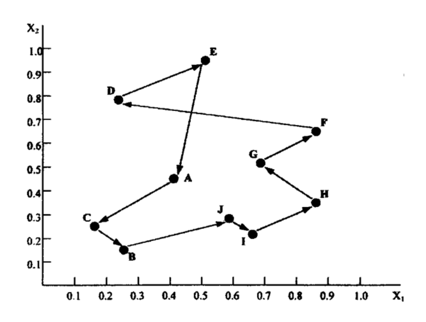
تطبيق عملي لآلة بولتزمان في مسألة البائع الجوال، استعمل فيه 100 هيئة بداية مختلفة، كل منها مع نصف الوحدات on، ومع $T_0= 20$ و$b = 60$ و$p = 70$، نتجت الجولات الصحيحة في 20 دوراً أو أقل لكل 100 هيئة أولية. في هذه التجارب، كان من النادر أن تغير الشبكة هيئتها مرة واحدة لتعطي جولة صحيحة مباشرة.

نموذجياً، وجدت الجولة الصحيحة بعد عشرة أدوار أو أقل. يتألف الدور من كل وحدة تحاول تغير حالتها. كانت جدولة التبريد $T^{new} = 0.9\ T^{old}$ بعد كل دور. وقد وجدت الجولات الخمس التالية بطول أقصر من 4:

| الطول | الجولة | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3.036575 | *H* | *I* | *J* | *B* | *C* | *A* | *E* | *D* | *F* | *G* |
| 3.713347 | *B* | *C* | *E* | *H* | *F* | *G* | *J* | *I* | *A* | *D* |
| 3.802492 | *C* | *D* | *E* | *I* | *G* | *F* | *A* | *H* | *J* | *B* |
| 3.973623 | *G* | *F* | *D* | *C* | *B* | *A* | *J* | *E* | *I* | *H* |
| 3.975433 | *I* | *G* | *B* | *C* | *E* | *D* | *H* | *F* | *A* | *J* |

والجولة المثلى الموجودة موضحة في الشكل (8.9).

استُخدمت معطيات أخرى بقيم أصغر للأوزان b وp ‏($b = 30$ , $p = 35$)، فوُجدت 100 جولة صحيحة أيضاً لكل 100 هيئة أولية. تغيرت الشبكة من جولة صحيحة إلى جولة صحيحة ثانية (أقصر) تقريباً 25% من هذه المسالك.




**الشكل 8.9:** أفضل جولة للبائع الجوال من آلة بولتزمان (100 هيئة أولية)

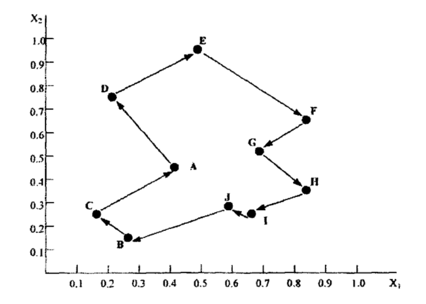
على أية حال ، لم تجد أي جولة أقصر مما وجد في المثال السابق. حتـــى باستعمال قيم أصغر(b = 6, p = 7) كانت الشبكة غير قادرة على إيجاد رحلات صحيحة (في 20 دوراً) لأية 100 هيئة أولية. من الملاحظ أن زيادة الأدوار غير مرغوب فيها وغير مساعدة لأن درجة الحرارة بعد 20 دوراً ستكون منخفضة تماماً.

طبعاً كما ذكر من قبل، حاول الكثيرون حل مشكلة البائع الجوال مثل Hopfield-Tank عام 1985[124] اللذين توصلا إلى مقدار عال من النجاح في إيجاد رحلات صحيحة؛ فقد أوجدا 16 جولة صحيحة من 20 هيئة بداية. تقريباً نصف المسالك أعطت واحداً من أقصر ممرين. الجولة الأفضل التـــي وجدت، والموضحة في الشكل (9.9) هي:

| الطول | الجولة | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 2.71 | *C* | *B* | *J* | *I* | *H* | *G* | *F* | *E* | *D* | *A* |

في حل Hopfield-Tank لمسألة البائع الجوال عام 1985[124]، كان لكل وحدة دليلان، الدليل الأول يشير إلى المدينة والدليل الثانـــي يشير إلى المكان في الجولة. يعطى تابع طاقة Hopfield-Tank لمسألة البائع الجوال كما يلي:

$$E = \frac{A}{2}\sum_{x}\sum_{i}\sum_{j\neq i}\upsilon_{xi}\upsilon_{xj} + \frac{B}{2}\sum_{i}\sum_{x}\sum_{y\neq x}\upsilon_{xi}\upsilon_{yi} + \frac{C}{2}[N - \sum_{x}\sum_{i}\upsilon_{xi}]^2 + \frac{D}{2}\sum_{x}\sum_{y\neq x}\sum_{i}d_{xy}\upsilon_{xi}(\upsilon_{y,i+1} + \upsilon_{y,i-1}) \quad (38.9)$$




**الشكل 9.9**: أفضل جولة للبائع إلجوال موجودة من قبل Hopfield-Tank

المعادلة التفاضلية لفعالية الوحدة $U_{X,I}$ تعطى بما يلي:

$$\frac{d}{dt}u_{X,I} = -\frac{u_{X,I}}{\tau} - A\sum_{j\neq I}\upsilon_{X,I} - B\sum_{y\neq X}\upsilon_{y,I} - C[N - \sum_{x}\sum_{i}\upsilon_{xi}] - D\sum_{y\neq X}d_{Xy}(\upsilon_{y,I+1} + \upsilon_{y,I-1}) \quad (39.9)$$

أعطيت إشارة الخرج بتطبيق تابع تفعيل sigmoid (بمجال بين الصفر والواحد)، حيث عبر عنه Hopfield-Tank كما يلي:

$$\upsilon_i = g(u_i) = 0.5[1 + \tanh(\alpha u_i)] \quad (40.9)$$

يمكن تلخيص الخوارزمية التـي استخدمها Hopfield-Tank بما يلي:

1. ضع التفعيلات الأولية لكل الوحدات، ضع القيمة الأولية لـ $\Delta t$ بقيمة صغيرة.

2. مادام شرط التوقف غير محقق كرر الخطوات من 3 إلى 6.
3. اختر وحدة عشوائياً.
4. غيّر تفعيل الوحدة المختارة.

$$\begin{aligned} u_{x,i}(new) = u_{x,i}(old) + \Delta t[-u_{x,i}(old) - A\sum_{j\neq i} \upsilon_{x,j} \\ - B\sum_{y\neq x} \upsilon_{y,i} - C[N - \sum_{x}\sum_{j} \upsilon_{xji}] \\ - D\sum_{y\neq x} d_{xy}(\upsilon_{y,i+1} + \upsilon_{y,i-1})] \end{aligned} \quad (41.9)$$

5. طبقّ تفعيل الخرج:

$$\upsilon_{xi} = 0.5[1 + \tanh(\alpha u_{xi})] \quad (42.9)$$

6. اختبر شرط التوقف.

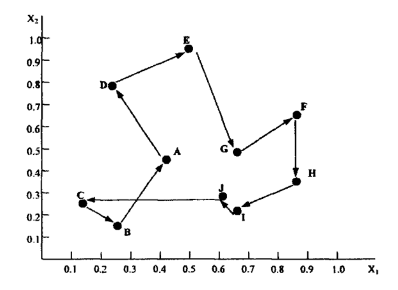
استعملَ Hopfield-Tank قيم الوسطاء التاليـــة في سعيهما لحل مسألة البائع الجوال: $A = B = 500, C = 200, D = 500, N = 15\ \alpha = 50$ اختيـــرت مستويات الفعالية الأولية (uxi) بحيث $\sum_{x}\sum_{i}\upsilon_{xi} = 10$ (التفعيل الكلي المرغوب به لجولة صحيحة). على أية حال كان هناك بعض الضجيج، لذا لم تبدأ كل الوحدات بنفس الفعالية( أو بإشارة خرج).

حاوَل Wilson وPawley عام 1988[122] الحصول على مقدار من النجاح بإجراء بعض التعديلات الطفيفة على خوارزمية Hopfield-Tank. فقد استعملا خطوة زمنية $\Delta t = 10^{-5}$، وأوقفا محاكاتهما عندما وجدا جولة صحيحة؛ أي جُمِّدت الشبكة، أو نفذا 1000 دور. اعُتبرت الوحدة on إذا كان تفعيل الوحدة أكبر من 0.9 وoff إذا كان أقل من0.1، وجُمِّدت الشبكة إذا كان تغير التفعيلات أقل من$10^{-35}$.

لقد أوجدا 15 رحلة صحيحة في 100 محاولة و1000 دور لكل محاولة (و45 حالة تجمد، و40 حالة عدم تقارب). بعض هذه الجولات الصحيحة (بما فيها الجولة الموضحة فـــي الشكل (9.9)) أعطى الاستمثال النتائج التالية:

| الجولة | | | | | | | | | | الطول |
|---|---|---|---|---|---|---|---|---|---|---|
| $A$ | $D$ | $E$ | $F$ | $G$ | $H$ | $J$ | $I$ | $B$ | $C$ | 2.71 |
| $A$ | $D$ | $E$ | $F$ | $G$ | $H$ | $I$ | $J$ | $B$ | $C$ | 2.71 |
| $A$ | $D$ | $E$ | $G$ | $F$ | $H$ | $I$ | $J$ | $C$ | $B$ | — |
| $A$ | $D$ | $B$ | $D$ | $E$ | $G$ | $F$ | $H$ | $J$ | $I$ | — |

يوضح الشكل (10.9) إحدى الجولات الفُضلى من قبل Pawley & Wilson.




**الشكل 10.9**: إحدى الجولات الفُضلى معطاة من قبل Pawley & Wilson.

حصل باحث آخر هو Szu عام 1988 [123] على نتائج جيدة باستعمال شبكة المعدلة، فقد استعمل تابع الطاقة التالي:

$$
\begin{aligned}
E = &\frac{A}{2}\sum_{x}\sum_{i}\sum_{j\neq i}\upsilon_{xi}\upsilon_{xj} \\
&+\frac{B}{2}\sum_{i}\sum_{x}\sum_{y\neq x}\upsilon_{xi}\upsilon_{yi} \\
&+\frac{C}{2}\{\sum_{x}[1-\sum_{i}\upsilon_{xi}]^{2}+\sum_{i}[1-\sum_{x}\upsilon_{xi}]^{2}\} \\
&+\frac{D}{2}\sum_{x}\sum_{y\neq x}\sum_{i}d_{xy}\upsilon_{xi}(\upsilon_{y,i+1}+\upsilon_{y,i-1})
\end{aligned}
\tag{42.9}
$$

حيث يلبـــي الحد الثالث في هذا التابع ضرورة كون وحدة واحدة on في كـــل سطـــر وكل عمود. واختار الوسطاء بالقيم التالية :

$$A = B = C = D = 1$$

بالإضافة إلى تحسين تابع الطاقة، استعمل Szu تفعيلات مستمرة لكن بإشارات خرج ثنائية؛ كان تابع الخرج هو تابع الخطوة الواحدية، بدلاً من تابع sigmoid التفاضلي الذي استعمله Hopfield-Tank. يسمـــى تابع الخطوة أيضاً تابـــع دخل/خرج McCullosh-Pitts (Takefuji عام 1992[125]).

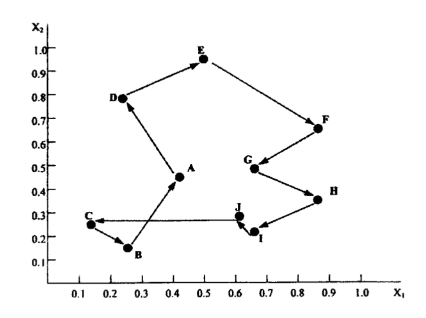
يمكن تلخيص خوارزمية الاستمثال السريعة لمسألة البائع الجوال بما يلي:

1. ضع المسافات الأولية بين المدن بقيم أولية، وضع قيمة $\Delta t = 10^{-5}$ .
2. نفذ الخطوات من 3 إلى 9 بعدد محدد من المرات (توليد عدد محدد من الأدوار).
3. ضع التفعيلات الأولية لكل الوحدات، استعمل قيماً عشوائية بين [0.0005+ ,0.0005 –]، زد تفعيل الوحدة $U_{11}$ بمقدار 0.005
4. كرر الخطوات من 5-8 $n^2$ مرة .
5. كرر الخطوات 6 إلى 7 لكل وحدة.
6. احسب كل الحدود لكل تغيّر في الفعالية.
7. حدّث الفعالية.
8. طبّق تابع الخرج الثنائي على كل وحدة.
9. اختبر هل الجولة صحيحة أم لا.

نشر Szu نتائجه عام 1988 [123] لخوارزمية استمثال سريعة تعطي 91 جولة صحيحة من 1000 مسلك(جولة مولدة). وكانت أفضل جولاته على الترتيب التالي:

| الطول | الجولة |
| --- | --- |
| 2.76693 | $A$ $B$ $C$ $J$ $I$ $H$ $G$ $F$ $E$ $D$ |

الموضحة في الشكل (11.9).




**الشكل11.9** : أفضل جولة للبائع الجوال من خوارزمية الاستمثال السريعة لـــ Szu

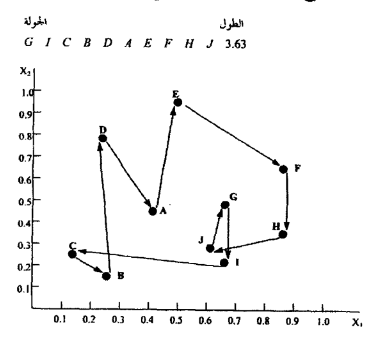
وقد وجدت الجولات التالية بطول أقل من 3.5.

| الطول | الجولة | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 3.3148 | *I* | *C* | *A* | *B* | *D* | *E* | *F* | *G* | *H* | *J* |
| 3.3306 | *E* | *D* | *F* | *H* | *I* | *J* | *G* | *B* | *C* | *A* |
| 3.3647 | *E* | *D* | *C* | *B* | *A* | *F* | *H* | *G* | *I* | *J* |
| 3.3679 | *D* | *B* | *C* | *I* | *J* | *H* | *F* | *G* | *E* | *A* |
| 3.3822 | *A* | *J* | *I* | *G* | *H* | *F* | *D* | *E* | *B* | *C* |
| 3.4345 | *B* | *C* | *I* | *J* | *H* | *G* | *E* | *D* | *F* | *A* |
| 3.4917 | *A* | *D* | *J* | *I* | *H* | *F* | *G* | *E* | *B* | *C* |

استخدمت آلة كوشي ـــ بولتزمان الهجينية أيضاً لحل مشكلة البائع الجوال، وقد وجد جولات بطول أقل من أربعة من أجل 9 هيئات أولية، ولكنها لم تجد أية جولة أقصر مما وجد باستخدام آلة بولتزمان لمفردها. فقد وجدت 4 جولات بطول أقل من 3.5 ووجدت واحدة فقط كانت هي الأفضل من بين الجولات المتولدة بواسطة آلة بولتزمان.

| الجولة | | | | | | | | | | الطول |
|---|---|---|---|---|---|---|---|---|---|---|
| J | G | E | D | A | B | C | I | H | F | 3.3341 |
| I | J | H | E | D | A | C | B | F | G | 3.3968 |
| I | J | D | A | B | C | E | G | F | H | 3.4649 |
| I | E | D | F | G | H | A | C | B | J | 3.4761 |
| J | F | E | A | C | B | I | H | G | D | 3.8840 |
| F | C | B | A | D | G | E | I | J | H | 3.8944 |
| H | J | A | I | B | C | F | D | E | G | 3.9045 |
| C | J | H | I | G | D | E | F | B | A | 3.9513 |
| F | H | D | J | I | A | C | B | G | E | 3.9592 |

استخدمت آلة كوشي لمفردها أيضاً في حل مشكلة البائع الجوال، في هذه الحالة أوجدت الشـــبكة 11 جولة بطول أقل من 4 من أجل 100 هيئة أولية عشوائية، وكانت أقصر جولة بطول 3.63 كما هو موضح في الشكل (12.9) وهي:




**الشكل 12.9:** أفضل جولة للبائع الجوال باستعمال آلة كوشي

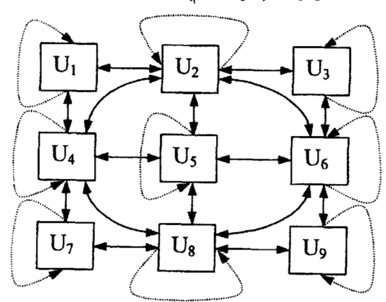
أخـــيراً ، بالطبع هناك تطبيقات أخرى عديدة لهذا الصنف من الشبكات لا يسعنا ذكرها مثل المشروع الأوروبـــي ANNIE؛ (Applicationa of Neural Networks for Industry)

(in Europe الـــذي رعاه برنامج الذكاء الصنعي الأوروبـــي ESPRIT في تشرين الثانـــي عام 1988، وذلك لتطوير ونشر تطبيقات الشبكات العصبونية الصنعية في أوربا.

وقد نشرت نتائج هذه الجهود المنصبة على مسألة جدولة الملاحة الجوية في مجلد يحمل لقب Project ANNIE Handbook من قبل دار نشر Croall & Mason عام 1991[126].

## 6.9 تمارين

1.9 لتكن لدينا شبكة آلة بولتزمان (بدون تعليم) التالية:




**الشكل 13.9**: شبكة آلة بولتزمان

جميع الوصلات لها وزن مرافق يساوي ‎-2، وكل وحدة لها وصلة ذاتية بوزن يساوي الواحد. افترض أن حالة الوحدة $U_2$ هي on وكل الوحدات الأخرى في حالة off. صف ماذا يحدث في كل من الحالات التالية :

1. $T=10$ و $U_9$ تحاول أن تكون on
2. $T=1$ و $U_9$ تحاول أن تكون on
3. $T=10$ و $U_6$ تحاول أن تكون on
4. $T=1$ و $U_6$ تحاول أن تكون on

2.9 استُعملت آلة بولتزمان لحل مسألة البائع الجوال. المسافات بين المدن المشكلة للرحلة كما يلي:

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 4 | 3 | 2 |
| B | 4 | 0 | 5 | 6 |
| C | 3 | 5 | 0 | 1 |
| D | 2 | 6 | 1 | 0 |

بافتراض أن الغرامة p = 20 والمكافأة b = 15 كشروط مقيدة (الأوزان)، وT=100. في الجولة CDBAC:

1. ما هي تفعيلات الوحدات، أي ما هي الوحدات التـــي ستكون on أو off؟.
2. ارسم مخطط الشبكة للجولة ، بوصلات ظاهرة في المخطط للوحدات الفعالة. (ضع كل الشروط المقيدة والمسافات بين الوحدات الفعالة)
3. احسب تغير الإجماع لكل وحدة فعالة (إذا حاولت أن تكون off).
4. في كل وحدة فعالة، احسب احتمال تغير حالتها.

3.9 استُعملت آلة بولتزمان لحل مسألة البائع الجوال للمدن التالية في الأماكن التالية:

| المدينة | المكان | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| A | $U_{A1}$ | $U_{A2}$ | $U_{A3}$ | $U_{A4}$ |
| B | $U_{B1}$ | $U_{B2}$ | $U_{B3}$ | $U_{B4}$ |
| C | $U_{C1}$ | $U_{C2}$ | $U_{C3}$ | $U_{C4}$ |
| D | $U_{D1}$ | $U_{D2}$ | $U_{D3}$ | $U_{D4}$ |

والمسافة بين المدن الأربع كما يلي:

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 6 | 8 | 5 |
| B | 6 | 0 | 10 | 5 |
| C | 8 | 10 | 0 | 5 |
| D | 5 | 5 | 5 | 0 |

باستعمال غرامة P = 20 ومكافأة B = 10 كشروط مقيدة.

1. احسب أوزان وصلات الوحدة $U_{C3}$.

2. حدد قيمة الإجماع للشبكة إذا كان للوحدات التفعيلات التالية:

| المكان | المدينة | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| A | 1 | 0 | 0 | 0 |
| B | 0 | 1 | 0 | 0 |
| C | 0 | 0 | 1 | 1 |
| D | 0 | 0 | 1 | 0 |

3. حدد قيمة الإجماع للشبكة إذا كان للوحدات التفعيلات التالية:

| المكان | المدينة | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| A | 1 | 0 | 0 | 0 |
| B | 0 | 1 | 0 | 0 |
| C | 0 | 0 | 1 | 0 |
| D | 0 | 0 | 0 | 1 |

4. أي الهيئات في الطلبات 2 و3 تحقق كل الشروط المقيدة لمسألة البائع الجوال؟.

5. ما هو التأثير في الإجماع (أوجد $\Delta C$) إذا عكس تفعيل الوحدة $U_{C3}$ في الشبكة للطلب 2؟ وفي الشبكة للطلب 3؟.

6. لكل حالة معتبرة في الطلب 6، أوجد احتمال قبول التغيـــر إذا كان $T = 10$، و$T = 1$.

4.9 في مسألة البائع الجوال، المسافة بين المدن معطاة بالمصفوفة التالية:

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 8 | 10 | 20 | 5 |
| B | 8 | 0 | 26 | 20 | 9 |
| C | 10 | 26 | 0 | 10 | 5 |
| D | 20 | 20 | 20 | 0 | 5 |
| E | 5 | 9 | 5 | 5 | 0 |

استعمل $p = 70$ و$b = 60$ و$T=100$.

1. ارسم الشبكة مع الوصلات والأوزان لتمثيل الشروط المقيدة لهذه المسألة (لكن بدون

مسافات) .

2. ارسم الشبكة مع الوصلات والأوزان لتمثيل المسافات لهذه المسألة (لكن بدون شروط مقيدة) .
3. في الرحلة التالية: BACEDB ما هي تفعيلات الوحدات؟هذا يعنـــي أن أي وحدة ستكون on أو offفي كل عمود.
4. ما هي قيمة تابع الإجماع C لهذا الترتيب ؟
5. احسب $\Delta C$ لكل وحدة حالياً هي on
6. لكل وحدة حالياً on ، احسب احتمال تغير حالاتها من on إلى off.

5.9 اكتب برنامجاً لأداء آلة بولتزمان بدون تعليم لحل مسألة البائع الجوال بين خمس مدن بالمسافات المعطاة كما يلي:

| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 8 | 10 | 20 | 5 |
| B | 8 | 0 | 26 | 20 | 9 |
| C | 10 | 26 | 0 | 10 | 5 |
| D | 20 | 20 | 10 | 0 | 5 |
| E | 5 | 9 | 5 | 5 | 0 |

إضافـــة للمسافات ستكون الأوزان اللازمـــة $b > 2d$ و$p>b$، حيث d تحسب من أجل $b = 60$ و$p = 70$. استعمل مولد أعداد عشوائية يعطي قيمة بين الصفر والواحد، يمكنك اختيار عشوائياً الوحدة التـــي تحاول تغير إجماعها بأخذ $r_1$ 5=i و$r_2$ 5=j، حيث $r_1$ و$r_2$ هما عددان عشوائيان مختلفان وi وj أرقام صحيحة مقطوعة الفاصلة العشرية وما بعدها. يمكن اعتبار المسافات ملفوفة وذلك بتعريف $jplus = j + 1 mod 5$، إذا كـــان $jplus = 6$ فإن $jplus = 1$ (وبالمثل لـــ jminus). بعد حساب احتمال قبول A، ولدٍّ عدداً عشوائياً (ثانية بين الصفر والواحد، بتوزيع طبيعي). إذا كان الرقم المولد أصغر من A غيّر تفعيل الوحدة (i , j) يمكن أن ينجز عكس التفعيل بواسطة:

$$x(i,j) = x(i,j) + 1 \bmod 2$$

ابدأ بـــ $T = 10$ وقلّل T بطريقة خطية حتـــى الصفر. نفّذ برنامجك مرات عديدة، ابدأ بوحدات أولية مختلفة on في كل مرة، وحاول أن تبدأ بدرجات حرارة مختلفة، و..الخ.

# الشبكات العصبونية الصنعية الذاتية النمو
# Self-Growing Artificial Neural Networks

هذا هو الفصل الأول من ثلاثة فصول متتالية تناقش نوعاً هاماً من بنـــى الشبكات العصبونية الصنعية. تعتبر أصناف الشبكات العصبونية الموصوفة في هذه الفصول واحدة ولكنها مختلفة بعضها عن بعض تماماً من وجهة النظر التالية: ينبـــى الصنف الأول ذاتياً، ويمتاز الصنف الثانـــي بنظام توصيل اختياري بين الطبقات، أما الصنف الثالث فيبنـــي نماذج احتمالية (stochastic) للوسط المحيط.

سنبحث في هذا الفصل عن تقريب جديد لبناء الشبكات العصبونية الأمامية التغذية، ستكون هذه الشبكات مبنية ذاتياً، حيث تقوم خوارزمية التدريب بإضافة طبقة مخفية أو أكثر بالتتالي خلال عملية التدريب بمعلم. يستمر هذا البناء الذاتـــي بإضافة عقد مخفية مع جميع وصلاتها حتـــى تصل قيمة الأخطاء الناتجة عن خوارزمية التدريب إلى مستويات مقبولة.

يستخدم أول نوعين من هذه الشبكات لمسائل التطبيق (mapping) العام، ويكون النوع الثالث مناسباً أكثر لمسائل التصنيف العام. جميع هذه الشبكات هي متعددة الطبقات وأمامية التغذية ومدربة باستعمال خوارزميات التدريب بمعلم.

## 1.10 تمهيد

إن الشبكات العصبونية الصنعية المدروسة عبر جميع الفصول السابقة كانت مبنية يدوياً، حيث لم يكن الحجم الأمثلي لهذه الشبكات معروفاً لكل مسألة معطاة على العموم، بل يجب تعيينه من خلال التجربة وعملية تقليل الأخطاء. وكما رأينا، كانت العملية أحياناً تنتهي بالإخفاق، مثلاً عند الوقوع في مشكلة الأصغر المحلي خلال التدريب، وكان من الضروري

عند تغير الوسط المحيط التفكير ببنية جديدة مختلفة لتتكيف مع هذا التغير.

وعلى الرغم مما أنجز من أبحاث وتطبيقات هامة ناجحة على الشبكات المتعددة الطبقات الأمامية التغذية لتعيين العدد الأمثلي لطبقاتها وعقدها المخفية، فما تزال هناك حاجة ماسة للبحث والتجريب بهذا الخصوص. وقد ذُلِّلت بعض المصاعب باستعمال الشبكات الذاتية النمو الموصوفة في هذا الفصل، وهذه الشبكات لا تعانـــي مطلقاً من مشاكل الأصغر المحلي.

كما ذكر من قبل، تقوم خوارزمية تدريب هذه الشبكات بإضافة عقد مخفية، عقدة تلو الأخرى، حتـــى يصل الخطأ إلى المستوى المقبول.

سندرس فيما يلي أربعة أنواع لشبكات النمو الذاتي؛سنستعرض أول شبكتين بالتفصيل مع بعض التطبيقات النموذجية، أما الشبكتان الأخيرتان فسندرسهما باختصار.

سنبدأ بوصف شبكة طاقة كولومب (Coulomb) المُخفَّضة التـــي سجلت اختراعاً، بعدئذ سنناقش شبكة الارتباط المتتابع (cascade)، وأخيراً سنستعرض عمل الشبكات البرجية والهرمية وخوارزمية الانطلاق (upstart).

## 2.10 شبكات طاقة كولومب المخفضة

**Reduced Coulomb Energy Networks(RCE)**

طور Cooper وReily وزملاؤهما شبكات طاقة كولومب المخفضة بين عام 1982 [202] وعام 1987[203]. عُلِّمت هذه الشبكات على تطبيق (mapping) شعاع ملامح الدخل **x** بقيم حقيقية من فراغ ذي بعد **n** إلى فراغ فئة الخرج C (من بين **c** فئة ممكنة) من خلال معادلة الطاقة التالية:

$$E = \mathbf{x} \rightarrow C \qquad (1.10)$$

حيث

$$C = E(\mathbf{x})$$

اشتق اسم طاقة كولومب المخفضة من علاقة الشبكة بالفيزياء، حيث يُعرَّف تابع الطاقة E (في الكهرباء الساكنة) بالاعتماد على مواقع ذاكرة الشبكة في فراغ ذي بعد **n**. للاستفادة من التشابه الفيزيائي، سيطبق الدخل E على الخرج C بواسطة التطبيق R، حيث $C = R^{-1} E$

أو R C = E. تعريف الطاقة، الذي سنهمله هنا، له أحواض ذاكرة "كولومب" بصغريات محلية فقط عند مواقع الذاكرة الموافقة (خلافاً لشبكة هوبفيلد).

الشبكة هي شبكة تغذية أمامية بوصلات مخفضة وبثلاث طبقات معالجة: طبقة الدخل مؤلفة من n خلية، وطبقة التطبيق الداخلية أو المتوسطة مؤلفة من m خلية، وطبقة الخرج مؤلفة من c خلية.

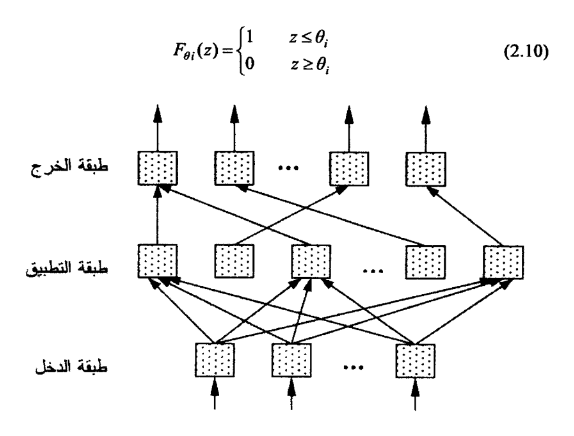
توافق الخلايا في طبقة الدخل قيم الملامح (السمات/المعالم) أو الصفات المميزة لنموذج الدخل، على حين توافق كل خلية من الخلايا في طبقة الخرج صفاً أو فئة للملامح المميزة المختلفة، وتقوم الخلايا في الطبقة الداخلية بتطبيق قيم ملامح الدخل إلى فئات الخرج.

تخصص الشبكة شعاع الدخل إلى فئة خاصة عندما تتنشط عقدة الخرج (تساوي الواحد). تسمّى أشعة ملامح الدخل التـــي تعطي استجابة خرج وحيدة بالتطبيقات غير الغامضة (unambiguous)، وتسمّى أشعة الدخل التـــي تعطي استجابات خرج متعددة أو لا تعطي استجابة خرج نهائياً بالتطبيقات الغامضة (ambiguous). على أية حال، تبقى التطبيقات الغامضة التـــي تعطي مخارج ما مفيدة إذا استطاعت أن تعطينا مؤشرات عن الفئات المشابهة جداً لشعاع معالم الدخل. وهذا ما سنركز عليه فيما بعد.

تتصل كل خلية في الطبقة المتوسطة اتصالاً كاملاً مع جميع خلايا طبقة الدخل بواسطة أوزان قابلة للتعديل. توجه خلايا الطبقة المتوسطة **وصلة خرجها الوحيدة على خلية طبقة خرج وحيدة فقط من خلال أوزان ثابتة.** لن يكون هناك وصلات جانبية ولا وصلات تغذية عكسية. يوضح الشكل (1.10) البنية الأساسية لشبكة طاقة كولومب. لاحظ أن خلايا الخرج يمكن، عموماً، أن يكون لها أكثر من وصلة دخل من خلايا الطبقة الداخلية.

كل خلية $i$ في الطبقة الداخلية تعرِّف منطقة تأثير (نفوذ) خاصة بها في فراغ النموذج مبنية على شعاع وزن دخلها $\mathbf{W_i}$ وعلى وسيط العتبة المرافق $\theta_i$. تُعرَّف هذه المنطقة من خلال تابع التفعيل $F_{\theta i}(d(\mathrm{w}_i,\mathrm{x}))$، حيث d مسافة (متري) معطاة مناسبة. مثلاً، يمكن أن تكون هذا المسافة ديكارتية، أو مسافة هامنغ، أو جداء داخلياً شعاعياً، أو قياساً آخر ما، و$\mathbf{X}$ قيم شعاع ملامح الدخل. يعرف تابع العتبة $F_{\theta i}(.)$ بما يلي:

$$F_{\theta i}(z)=\begin{cases}1 & z\leq\theta_i\\0 & z\geq\theta_i\end{cases} \tag{2.10}$$




**الشكل 1.10**: شبكة طاقة كولومب المخفضة

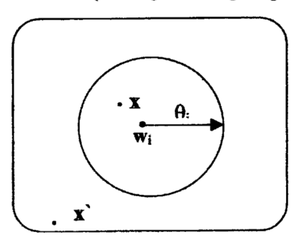
إذا كانت d مسافة ديكارتية في فراغ حقيقي القيمة ذي بعد n، فإن منطقة التأثير ستكون كرة بنصف قطر $\theta_i$ مركزها متوضع عند $w_i$. بعبارة مختصرة، تعرَّف منطقة تأثير الخلية $i$ في طبقة التطبيق بثلاث حقائق:

1. تعين نقطة توضع المركز في فراغ ذي بعد n بواسطة شعاع أوزان الخلية $\mathbf{w_i}$.
2. يعين حجم (نصف القطر) المنطقة بوسيط العتبة $\theta_i$.
3. يعين شكلها الهندسي (topology /geometry) بواسطة المتري d.

يوضح الشكل (2.10) منطقة النفوذ في فراغ الملامح الثنائي البعد بالقياس الديكارتي. نلاحظ في هذا الشكل أن النموذج **x** يقع ضمن منطقة النفوذ، أما النموذج **x`** يقع خارج المنطقة.

تعتبر خلايا الطبقة الداخلية فعالة بإعطاء قيمة خرج تساوي الواحد، على حين تكون المسافة بين قيمة ملمح الدخل **x** والأوزان الموافقة $\mathbf{w_i}$ للخلية أقل من قيمة العتبة $\theta_i$ (هذا

يعنـــي وقوع نموذج الدخل ضمن منطقة النفوذ للخلية).




**الشكل 2.10**: منطقة نفوذ الخلية i من طبقة التطبيق

وهكذا بالعودة إلى الشكل السابق نجد أنه عندما يقدم النموذج **x** إلى طبقة الدخل، تصبح الخلية *i* في الطبقة الداخلية فعالة ولكن لن تكون نفس الحالة (ليست فعالة) عندما يقدم النموذج **x`** إلى طبقة الدخل (لأن النموذج**x`** يقع خارج منطقة النفوذ).

ترسل الخلية الفعالة *i* خرجها إلى وحدة خرج واحدة فقط. وتسلك خلايا طبقة الخرج سلوك بوابة OR المنطقية، حيث تصبح فعالة إذا كان خطاً واحداً على الأقل من خطوط دخلها فعالاً. توضع الأوزان على الوصلات بين خلايا الطبقة الداخلية وخلايا طبقة الخرج بقيمة ثابتة مساوية للواحد.

يعتبر عمل شبكة طاقة كولومب المخفضة صحيحاً عندما تعطي مخارج غير غامضة لكل نموذج دخل **x**، وعندما تقوم المخارج بتصنيف صحيح لنماذج الدخل أيضاً. ينجز التصنيف الصحيح من خلال التدريب بمعلم لشبكة طاقة كولومب المخفضة، ويبدأ الإجراء بشبكة مركبة جزئياً من n عقدة دخل و c عقدة فئة خرج، حيث n عدد قيم ملامح شعاع الدخـــل و c عدد الفئات في الخرج. ليس من الضروري البدء في التدريب بأية طبقة داخلية أو عقدة طبقة داخلية متصلة، حيث تضاف حسب الحاجة خلال عملية التدريب كما هو مشروح فيما يلي.

## 3.10 تدريب شبكات طاقة كولومب المخفضة

**Training RCE Networks**

كما ذكرنا من قبل، يبدأ تدريب شبكات طاقة كولومب المخفضة بشبكة غير كاملة، أي شبكة بدون خلايا طبقة داخلية. سنفترض أن الشبكة مؤلفة من n خلية دخل لملاءمة بُعد أشعة نماذج التدريب و c خلية خرج، واحدة لكل فئة معروفة.

يُستعمل نموذج التدريب بمعلم حيث تختار النماذج من مناطق الفئات المختلفة كيفياً وتقدم إلى طبقة الدخل. إذا كانت استجابة الخرج صحيحة لنموذج الدخل لا يحدث أي فعل أو تعليم. أما إذا كانت خلية الخرج off (صفر) وأصبحت on (واحد) فستولد إشارة الخطأ (+1) وتعاد إلى الطبقة الداخلية (إشارة المعلم). وإذا كانت إشارة الخرج on (واحد) وأصبحت off (صفر)، تعاد إشارة الخطأ (−1) إلى الطبقة الداخلية. تستعمل إشارات الخطأ هذه في النظام لتدريب الشبكة.

يتألف التدريب من خطوتين:

1. إيداع خلية جديدة في الطبقة الداخلية
2. تعديل قيمة عتبة الخلايا الموجودة من قبل.

تودع خلية الطبقة الداخلية الجديدة في الشبكة بوصل دخل هذه الخلية إلى كل عقد الدخل ووصل خرجها الوحيد مع عقدة خرج مناسبة، ووضع قيم شعاع الوزن للخلية وكذلك وضع قيمة العتبة الأولية للخلية. تعدل عتبة الخلية بتخفيضها بكمية ثابتة $\Delta\theta$.

مثلاً، عندما تستقبل إشارة الخطأ +1 من وحدة الخرج رقم k تودع خلية طبقة داخلية جديدة $i$ (تضاف إلى الشبكة) في الشبكة وتوصل إلى خلية الخرج رقم k. تقوم وصلات الدخل القادمة من كل عقدة دخل بتفعيل الخلية الجديدة المودعة، وتوضع الأوزان القابلة للتعديل $\mathbf{w_i}$ على هذه الوصلات بحيث تكون مساوية لشعاع نموذج الدخل $\mathbf{x}$. وهذا يضمن تفعيل خلية الخرج k عندما يقدم نموذج الدخل الحالي $\mathbf{x}$ إلى الشبكة. توضع عتبة الخلية الجديدة وفقاً لـــ:

$$\theta_i = \max\{\theta_{\max}, \theta_{opp} \qquad (3.10)$$

حيث $\theta_{opp}$ هي المسافة إلى مركز حقل التأثير الأقرب لأي خلية من صف نموذج مختلف عن صف نموذج الدخل **x**، وبعبارة أخرى، $\theta_{opp}$ هي المسافة الصغرى بين شعاع الوزن **$w_i$** وشعاع الوزن **$W_j$** للخلية الأقرب من الفئة المختلفة، و $\theta_{\max}$ أكبر قيمة لأي عتبة مفروضة كقيمة أولية من قبل المستعمل. تضمن الخلية الجديدة المودعة أن خلية الخرج الصحيح ستنشط فيما إذا قدم نموذج الدخل المرافق، لتصحيح حالة إشارة الخطأ 1+.

عندما تستقبل الطبقة الداخلية إشارة الخطأ 1- الواردة (إشارة المعلم) من قبل خلية الخرج رقم **k**، تُخفَّض قيمة العتبة $\theta_i$ بمقدار $\Delta\theta$ لكل خلايا الطبقة الداخلية الفعالة المتصلة بالخلية رقم **k**. على الأقل تكون بعض الخلايا المتصلة بالخلية رقم **k** فعالة بوجه خاطئ. يقلل تخفيض قيمة وسيط العتبة حجم منطقة النفوذ للخلايا الداخلية المتصلة بخلية الخرج رقم **k**، وهذا بدوره يقلل احتمال اختيار الفئة الخطأ لنموذج الدخل المسبب للخطأ. يعطى ملخص إشارات التدريب هذه بالجدول (1.10) التالي:

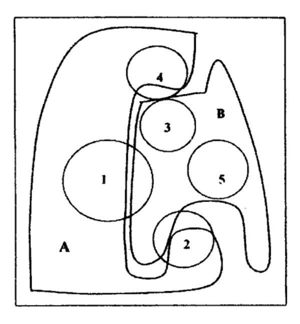
**الجدول 1.10** إجراءات التعليم وفقاً لأنواع إشارات الخطأ ±1

| إشارة الخطأ من خلية الخرج رقم k | الفعل المنفذ |
|---|---|
| 1+ | إبداع خلية طبقة داخلية ووصلها إلى وحدة الخرج رقم k. |
| 1- | تقليل قيم العتبة لكل خلايا الطبقة الداخلية الفعالة المتصلة مع خلية الرخج رقم k. |
| 0 | لا تغير في أي خلية. |

يوضح الشكل (3.10) مناطق نفوذ خمس خلايا طبقة داخلية(الدوائر المرقمة من 1 حتى خمسة) عند بداية التدريب. أتت نماذج التدريب من منطقتين لفئتين مختلفتين هما A وB المشاهدتين مع مناطق النفوذ للخلايا الخمس. نلاحظ عند هذه المرحلة من التدريب أن خلايا الطبقة المخفية الأولى والثانية ستعطي استجابات خاطئة (غامضة) عندما تقدم نماذج الفئة B المتوضعة ضمن منطقة التأثير لهذه الخلايا، وستعطى الاستجابات الصحيحة (غير

الغامضة) للنماذج المتوضعة ضمن منطقة تأثير الخلايا الثالثة والرابعة والخامسة.

لاحظ أن منطقة تأثير الخلية 3 أصغر من الخلايا الأخرى نتيجة تخفيض قيمة العتبة خلال عملية التدريب. ستعطي الخلية 4 نتيجة غير صحيحة إذا قدم نموذج ضجيجي متوضع بين المنطقتين A وB. هذه المشكلة يمكن أن تحل من خلال استعمال نماذج دخل ضجيجية خاصة في تدريب الشبكة كما سيوصف فيما بعد.

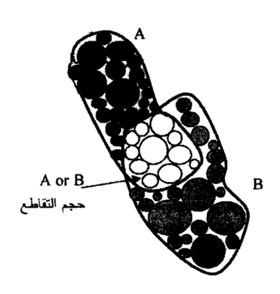



**الشكل 3.10**: مناطق نفوذ خمس خلايا طبقة داخلية عند بداية التدريب

إذا كانت مجموعة التدريب محدودة أو إذا كان اختيار قيم العتبة الأولية سيئاً لوحدات الطبقة الداخلية فإن مناطق الفئات لا يمكن أن تغطى بمناطق نفوذ الخلايا خلال التدريب، وهذا سيؤدي إلى تصانيف خاطئة للنماذج التـــي لم ترها الشبكة خلال التدريب. يوضح الشكل (4.10) مثالاً عن مناطق النفوذ المغطاة جزئياً لمناطق الفئتين A وB. حيث دربت الشبكة على النماذج من الصفين A وB.

بافتراض إعطاء الشبكة نماذج تمثيلية كافية وعدداً كافياً من الخلايا الداخلية المودعة، عندئذ ستكون الشبكة قادرة على تصنيف نماذج من فئات عديدة حتـــى وإن كانت مناطق الفئات معقدة ومتداخلة بما في ذلك مناطق الفصل غير الخطية بفراغ ذي بعد n. لتعلم

الفئات، قد نحتاج إلى مناطق نفوذ عدة خلايا لتغطية كل فئة، وخاصة إذا كانت المناطق معقدة جداً. إن نماذج التدريب الضجيجية هي النماذج المختارة قرب حدود المناطق، ولكنها تقع خارج مناطق الفئات. إذ يمكنها أن تساعد على جعل حدود مناطق الفئات حادة أكثر، وذلك لأن عتبات خلايا الطبقة الداخلية ستخفض بواسطة النماذج الضجيجية، على حين لن تكون هناك خلايا مودعة من هذه النماذج.

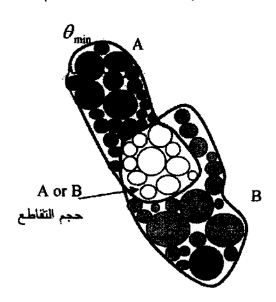



**الشكل 4.10**: مناطق النفوذ المغطاة جزئياً لمناطق فئات A وB

يتطلب تحضير نماذج التدريب الضجيجية بعض المعرفة بحدود الفئات. وهذا قد يكون من الصعب تحقيقه في بعض التطبيقات. لتلخيص ما قيل من أفكار، نلاحظ أن أي نموذج دخل يقع ضمن منطقة نفوذ الخلية سيفعل الخلية. وإذا وقع النموذج ضمن منطقة نفوذ عدة خلايا متشابكة فإن جميع هذه الخلايا ستنشط. وإذا كان بعض هذه الخلايا الفعالة يمثل فئات مختلفة عندها ستنشط خلايا خرج عديدة معطيةً تصنيفاً غامضاً. وينتج لدينا نتائج تصنيف غير غامضة عندما تكون كل خلايا الطبقة الداخلية الفعالة متصلة مع نفس خلية الخرج.

ومع أن التصانيف الغامضة ستكون خاطئة، فإن هناك بعض الحالات التـــي يكون فيها هذا التصنيف مفيداً لنوع من قياس الأرجحية لنوع الفئة. مثلاً عندما تكون مناطق الفئة متداخلة فيما بينها، فإنه من المفيد معرفة تقدير فئات النماذج عندما تقع في منطقة التداخل

كما هو موضح في الشكل (5.10). حيث يكون حجم التقاطع معرفاً بواسطة مناطق التداخل غير المفصولة لكلا الفئتين A وB. في النسخ المتطورة لشبكات طاقة كولومب المخفضة تنفّذ طريقة تقدير احتمالات حدوث للنماذج الواقعة ضمن حجم التقاطع لكل فئة. في هذه الحالة تستعمل العتبة الصغرى $\theta_{min}$ لتعريف الخلايا الاحتمالية التـــي يمكن أن تستعمل لأداء تصنيف هذه الصفوف الغير منفصلة (Scofield وزملاؤه عام 1988 [204]).




**الشكل 5.10**: تطبق المناطق المفصولة تعينياً، ويطبق حجم التقاطع (A or B) احتمالياً

## 1.3.10 تعلم الفئة ديناميكياً Dynamic category Learning

يمكن إجراء نوع من تعلم الفئة الديناميكي في شبكات طاقة كولومب المخفضة، وذلك لأن الفئات الجديدة يمكن أن تعلّم دون إعادة تدريب كاملة على النماذج القديمة. ينجز هذا بسهولة بواسطة تقديم نماذج الصف الجديد وتدريب النظام على هذه النماذج مع كل نماذج الصف المجاورة. تلزم نماذج الصف المجاورة للمساعدة على توفير شكل حاد لحدود مناطق الفئات وتوفير مناطق نفوذ غير متداخلة بين الصفوف الموجودة والصفوف المضافة حديثاً. هذه المفاهيم موضحة في الشكل (6.10)، حيث دربت الشبكة على تصنيف النماذج من فئة جديدة C بعد التدريب على تصنيف النماذج من الفئات B وA.

تمتاز شبكات طاقة كولومب المخفضة بالخواص التالية التـــي تفتقر إليها الشبكات

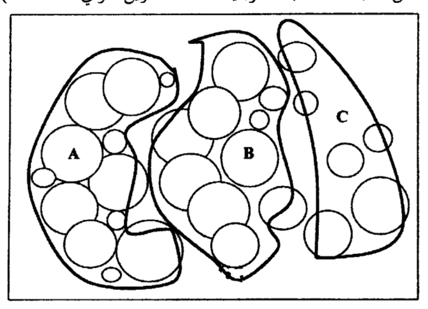
الأخرى:

ــ نمو ذاتـــي بحجم مناسب

ــ لا وجود لمشكلة الأصغر المحلي

ــ يلزم حسابات قليلة، وهذا ما يجعل تطبيقات الزمن الحقيقي ممكنة في بعض الشروط

ــ فعالية تخزين عالية، لأنه يلزم عقدة داخلية واحدة لتخزين أي نموذج دخل (استدعاء، حيث إن بعض الشبكات كشبكة هوبفيلد لها سعة تخزين حوالي 14% فقط).




**الشكل 6.10**: تعلم فئة جديدة C بعد تعلم الفئات A وB

ــ توصيل منخفض، حيث يكون عدد الوصلات تقريباً خطياً مع عدد العقد N بالمقارنة مع $N^2$ من أجل الشبكات الأخرى.

ــ قابلية تعليم سريعة وذلك من خلال بضع أمثلة تدريب فقط، والتعلم الديناميكي بدون إعادة تعلم كل الفئات الأخرى عندما يلزم تعلم وحدات جديدة.

هناك بالمقابل الخواص التالية التـــي تفتقدها هذه الشبكات:

ــ لا تعمم جيداً عادةً.

ــ تتطلب عقدة خرج فصل لكل فئة.

ــ يمكن أن تنمو نمواً ضخماً جداً في بعض المسائل.

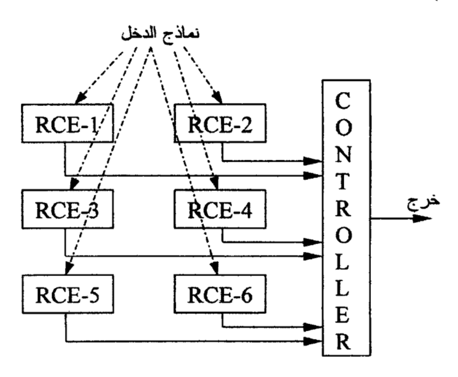
ليس بإمكانها إنجاز مهام توابع تطبيق (mapping) عامة.

## 4.10 شبكات طاقة كولومب المخفضة المتعددة والمتتالية

## Multiple and Cascaded RCE networks

استعملت شبكات طاقة كولومب المخفضة في تركيبات عنصر التحكم لحل بعض مسائل تصنيف النماذج المعقدة. مثلاً، يمكن أن يتألف الدخل من مجموعات الملامح المتعددة المتولدة بواسطة حساسات مختلفة أو أجزاء (أقسام) في فراغ الملمح كما هو مقترح بواسطة المواضيع العامة في الخواص المقاسة. يمكن أن يقدم التجزيء أيضاً طبيعياً نتيجة لملامح جديدة مكتشفة ومضافة إلى النظام مؤخراً.

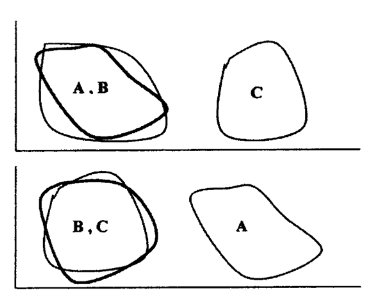
عندما يكون خرج الشبكة الوحيد غامضاً في بعض المسائل، قد توجد بعض الحلول غير الغامضة من خلال نوع من الارتباط أو خطة انتخاب (Reily عام 1987[203]). في هذه الحالة، يمكن أن توصل الشبكة في بنية متتالية لتركيب عنصر التحكم كما هو موضح في الشكل (7.10).




الشكل 7.10 : شبكات طاقة كولومب المخفضة بانتخاب الأكثرية

إن استعمال الشبكات المتعددة يحسن أحياناً دقة التصنيف في بعض التطبيقات. مثلاً،

عندما لا تستطيع الشبكة الأولى التمييز بين صفين A وB جيداً، ولا تستطيع الشبكة الثانية التمييز بين الصفين B وC، فإن استخدام الشبكتين معاً كما هـــو موضح في الشكل (8.10) يمكن أن يعطي حلاً لتصنيف جيد. لقد استعملت شبكات طاقة كولومب المخفضة على نطاق واسعٍ في تطبيقات مختلفة سنناقشها في نهاية الفصل.




الشكل 8.10: استعمال شبكات متعددة بانتخاب الأكثرية

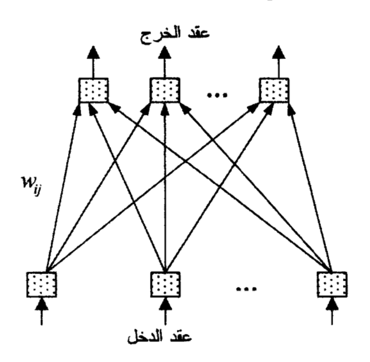
## 5.10 تدريب شبكات الارتباط المتتابع

## Training Cascade correlation networks(CNN)

اقترح Fahlman & Lebiere شبكات الارتباط المتتابع عام 1990[205] وهي نوع آخر من الشبكات الذاتية النمو خلال عملية التدريب.

الشبكة ككل هي شبكة تغذية أمامية بعقد دخل وخرج وطبقة مخفية. استعمل التدريب بمعلم للقيام بالتركيب المتزايد حتـــى الوصول إلى حجم الشبكة الأصغري اللازم للحصول على مستوى الخطأ المقبول. كما في شبكات طاقة كولومب المخفضة، تبدأ شبكة الارتباط المتتابع بعدد معين سلفاً من عقد الدخل والخرج وبشكل أولي كما هو متطلب للمسألة المعالجة.

توصل عقد الدخل كاملة إلى عقد الخرج من خلال أوزان قابلة للتعديل، بالإضافة إلى دخل الانحياز الثابت بقيمة 1+ دوماً. عند بداية التدريب لا يكون هناك أية عقدة طبقة مخفية ولا حتـــى طبقة مخفية، كما هو موضح في الشكل (9.10). تضاف عقد الطبقة المخفية إلى الشبكة عقدة تلو الأخرى في كل خطوة زمنية.




**الشكل 9.10**: شبكة الارتباط المتتابع الابتدائية قبل إضافة عقد مخفية

يمكن أن تدرب الشبكة باستعمال أي قاعدة تدريب بمعلم مثل البيرسبترون، أو الانتشار الخلفي، أو قاعدة دلتا البسيطة، أو أية طريقة أخرى مناسبة لأنه يجري تعديل أوزان طبقة واحدة فقط في لحظة ما خلال عملية التدريب.

قد تكون توابع التفعيل خطية أو غير خطية (تابع **simoid**، أو الظل القطعي، أو غير ذلك)، وقد تكون مختلطة (تابع جيبـــي، تابع **sigmoid**، تابع غوص)، أو قد تكون جميع هذه التوابع.

ينفّذ التدريب على شبكة بطبقة واحدة ابتدائية $w_{ij}$ حتـــى الوصول إلى عدم تحقيق أي تحسين في الخطأ (أصغر خطأ ممكن). إذا كان الخطأ مقبولاً يوقف التدريب، وإلا تضاف عقدة مخفية واحدة إلى الشبكة بوصل مداخلها إلى خرج كل عقد الدخل بأوزان $u_{1i}$ معدلة

(تشكيل طبقة مخفية جديدة) ولا يوصل خرج هذه العقدة المضافة إلى عقد أخرى في هذه اللحظة. بعدئذ تعدل الأوزان $u_{1i}$ على وصلات الدخل لجعل الارتباط C أعظمياً بين خرج العقدة المضافة والخطأ المتبقي من عقد الخرج، حيث

$$C = \sum_{o} \left| \sum_{p} \left( V_p - \overline{V} \right) \left( E_{p,o} - \overline{E}_o \right) \right| \qquad (4.10)$$

تتغير الأدلة p وo عبر كل عقد الخرج O ونماذج التدريب P على الترتيب، $V_p$ مخارج العقدة المضافة لنموذج الدخل $\mathbf{x}^p$، و $E_{p,o}$ الأخطاء المتبقية من عقد الخرج، حيث الخطأ المتبقي من عقد الخرج هو الفرق بين الخرج المنشود والمحسوب مضروباً بمشتق تابع تفعيل عقدة الخرج، وهذه هي الكمية التـــي ستنتشر عكسياً عند استعمال خوارزمية تعليم الانتشار الخلفي، حيث

$$E_{p,o} = \left[ \left( y_o^p - t_o^p \right) \right] y_o'^p \qquad (5.10)$$

$y_o^p$ الخرج المنشود للوحدة o في شعاع الدخل $\mathbf{x}^p$، و $t_o^p$ الخرج المحسوب للوحدة o في شعاع الدخل $\mathbf{x}^p$ و $y_o'^p$ مشتق الخرج المحسوب للوحدة o في شعاع الدخل $\mathbf{x}^p$، و $\overline{V}$ $\overline{E}_o$، متوسطات كل من $V_p$ و $E_{p,o}$ على الترتيب:

$$\overline{E}_o = \frac{1}{P} \sum_{p=1}^{P} E_{p,o} \qquad , \overline{V} = \frac{1}{P} \sum_{p=1}^{P} V_p \qquad (6.10)$$

يكون الارتباط أعظمياً (فعلياً هو القيمة المطلقة للتبايـــن المشترك) بإيجاد المشتق الجزئي لـــ $C$ بالنسبة لكل الأوزان الجديدة $\partial C / \partial u_i$ بأسلوب مشابه لإجراء قاعدة دلتا. بعد الحصول على الارتباط الأعظمي تجمّد قيم أوزان دخل العقدة (حتـــى نهاية التدريب) $u_{1i}$، ثم يوصل خرجها مع كل وحدات الخرج من خلال أوزان معدلة $v_{i1}$، حيث تعدل هذه الأوزان كجزء من كل أوزان وحدات الخرج $w_{ij}$ باستعمال قاعدة التدريب المعتمدة.

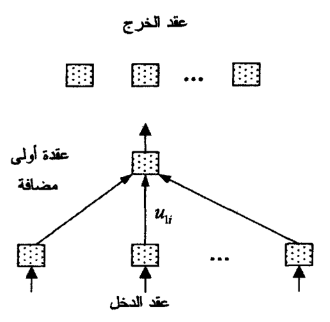
إذا بقي الخطأ غير مقبول تضاف عقدة طبقة مخفية أخرى إلى الشبكة، وتوصل بأوزان معدلة $u_{2i}$ مع كل عقد الدخل أولاً بالإضافة إلى العقدة المخفية المضافة من قبل عبر الوزن

$t_{21}$. في الواقع هذا سيضيف عقدة مفردة جديدة في طبقة مخفية ثانية جديدة. ينفّذ التدريب بنفس الإجراءات كما في عملية إضافة العقدة المخفية الأولى؛ أي تعديل أوزان دخل العقدة $u_{2i}$ مع الوزن $t_{21}$ لجعل الارتباط C أعظمياً، ومن ثم تجميد قيم هذه الأوزان (حتى نهاية التدريب).

أخيراً تدَّرب أوزان خرج العقدة الجديدة $v_{i2}$ مع باقي أوزان عقد الخرج $w_{ij}$، وتستمر هذه العملية حتى يصل الخطأ إلى مستوى مقبول عبر كل مجموعة التدريب. توضح الأشكال (10.10) حتى (13.10) مرحلتين متتابعتين حيث أضيفت عقدتان مخفيتان ووُصلتا إلى عقد الخرج.

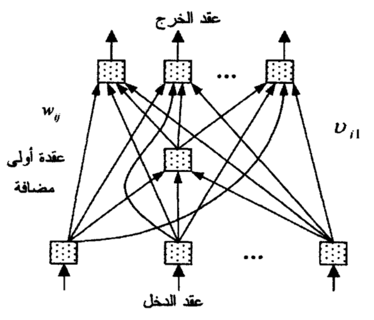
يمكن تلخيص خوارزمية تدريب الشبكة بما يلي، **(مع ملاحظة أن الأوزان المعدلة في كل مرحلة فقط ستظهر على الأشكال وباقي الأوزان غير الظاهرة ستكون مجمدة )**:

1. إيجاد شبكة ابتدائية بالعدد اللازم من عقد الدخل والخرج وموصلة بالكامل ووضع القيم الأولية لجميع الأوزان $w_{ij}$، كما هو موضح في الشكل (9.10).
2. تدريب الشبكة باستعمال قاعدة التدريب المختارة حتى يصل مستوى الخطأ إلى قيمته الصغرى، إذا وصل الخطأ إلى المستوى المقبول توقف وإلا اذهب للخطوة الثالثة.
3. إضافة عقدة طبقة مخفية مع وصلات دخلها فقط (بدون وصلة خرج) من مخارج كل وحدات الدخل وإعطاء الأوزان $u_{1i}$ قيماً أولية صغيرة، كما هو موضح في الشكل (10.10).
4. تعديل الأوزان $u_{1i}$ فقط (أوزان وصلات الدخل) للحصول على ارتباط أعظمي عبر مجموعة التدريب ككل، ومن ثم تجميد قيم هذه الأوزان حتى نهاية التدريب (جميع أوزان دخل العقدة الجديدة) بعد تقاربها.
5. وصل مخرج العقدة الجديدة إلى كل وحدات الخرج من خلال الوزن $v_{i1}$ وتعديل أوزان كل وحدات الخرج بما في ذلك الوزن $v_{i1}$ لتقليل الخطأ عبر كل نماذج التدريب، كما هو موضح في الشكل (11.10) .



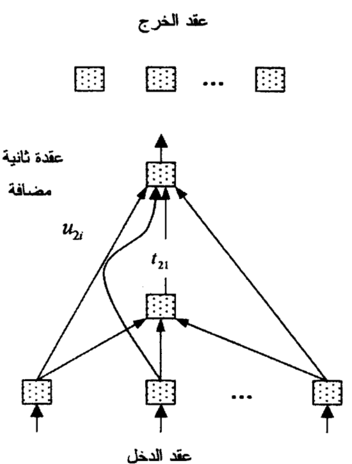

**الشكل 10.10**: شبكة الارتباط المتتابع بعد إضافة عقدة طبقة مخفية واحدة وبدون وصلة خرج

إذا وصل الخطأ إلى مستوى مقبول توقف، وإلا أضفْ عقدة مخفية ثانية (في طبقة مخفية جديدة) وَصِلْ مداخلها عبر الأوزان $u_{2i}$ إلى مخارج جميع عقد الدخل ومخارج جميع العقد المخفية السابقة عبر الوزن $t_{21}$، كما هو موضح في الشكل (12.10).



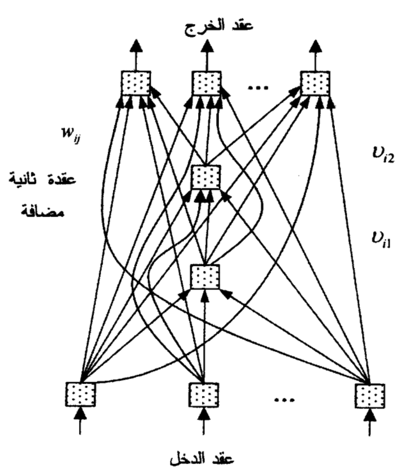

**الشكل 11.10**: شبكة الارتباط المتتابع بعد إضافة عقدة طبقة مخفية واحدة مع وصلة خرج.

6. تعديل الأوزان $u_{2i}$ والوزن $t_{21}$ فقط (أوزان وصلات الدخل) للحصول على ارتباط أعظمي عبر مجموعة التدريب ككل ومن ثم تجميد قيم الأوزان حتـــى نهاية التدريب (جميع أوزان دخل العقدة الجديدة) بعد تقاربها.




**الشكل 12.10**: شبكة الارتباط التصاعدي بعد إضافة عقدة طبقة مخفية ثانية وبدون وصلة خرج.

7. وصل مخرج العقدة الجديدة إلى كل وحدات الخرج من خلال الوزن $v_{i2}$ وتعديل أوزان كل وحدات الخرج بما في ذلك الوزن $v_{i1}$ و $v_{i2}$ لتقليل الخطأ عبر كل نماذج التدريب، كما هو موضح في الشكل (13.10) حتـــى الوصول إلى مستوى خطأ مقبول أو إلى نهاية الزمن المخصص للحساب.




**الشكل 13.10**: شبكة الارتباط التصاعدي بعد إضافة عقدتين (طبقتين) مخفيتين مع وصلة خرج.

يمكن إجراء تعديلات كثيرة على خوارزمية التدريب السابقة. مثلاً، عوضاً عن جعل الارتباط أعظمياً لعقدة مخفية مضافة واحدة، يمكن استخدام حوض من العقد، يكون فيه لكل عقدة مجموعة مختلفة من قيم الأوزان الأولية. يمكن أن يدرب حوض العقد ككل على التوازي لأن كل الوحدات تستعمل نفس مجموعة التدريب وتراقب نفس الخطأ المتبقي ولا تتقاطع فيما بينها. عندما لا يمكن أن يقلل الخطأ أكثر من ذلك، يجري اختيار وإضافة العقدة ذات الارتباط الأعظمي إلى الشبكة. يقلل هذا التقريب فرصة إضافة وحدة غير مفيدة (البقاء في الأصغر المحلي). لتحقيق الآلة (الحسابات) التفرعية، نستطيع تسريع التدريب لأن عدة أجزاء من فراغ الوزن يمكن أن تستكشف لحظياً (في آن واحد).

يمكن تحقيق تعديل آخر باستعمال تقليل الخطأ القياسي عوضاً عن تعظيم الارتباط كتابع التدريب الموضوعي. وهذا يتطلب بعض التعديلات في عملية نمو الشبكة كما وَصَف

Littman عام 1992 [206]. في هذه الحالة تدرَّب كل الوحدات المتتابعة لتقريب الخرج المنشود وكل وحدة مضافة حديثاً تصبح خرج الشبكة، حيث تعمل الوحدات المضافة قبلها كدخل فعلي للوحدة الجديدة.

تمتاز شبكة الارتباط المتتابع على الكثير من الشبكات المتعددة الطبقات الأمامية التغذية، بعدد من المزايا، منها:

+ ليس هناك حاجة لتخمين بنية الشبكة سلفاً، لأن خوارزمية التدريب تبنـــي الشبكةَ آلياً بحجم صغير يقابل معيار خطأ معين.
+ تعليم أسرع، مقارنة مع طرق أخرى مثل الانتشار الخلفي، تحل عملية التدريب المسألة تزايدياً مع عدم تعارض التقاطعات فيما بين العقد.
+ يمكن أن تبنـــى شبكات ضخمة بكواشف ملامح درجة عالية بدون أن يسبب ذلك إبطاء في التدريب كما وجد في الانتشار الخلفي عند استعمال أكثر من طبقتين مخفيتين.
+ يمكن القيام بتعليم معلومات جديدة دون القيام بإعادة تعليم كاملة، فقط يمكن أن يعاد تدريب أوزان الخرج.
+ حسابات التدريب أبسط من غيرها، لأنه يجري تدريب طبقة واحدة للأوزان فقط في كل لحظة، ويمكن أن تخبأ النتائج خلال عملية التدريب.
+ ليس هناك حاجة إلى إرسال إشارات الخطأ عكسياً من خلال الشبكة، إذ يتطلب التدريب إرسال إشارة في اتجاه واحد فقط وذلك تبسيطاً للحسابات.
+ لما كانت الوحدات المضافة لا تتقاطع فيما بينها، فإن اتصالاً محدداً يجعل البنية ميالة إلى العمل التفرعي.

وبوجه مشابه للشبكات المتعددة الطبقات الأمامية التغذية المدربة بخوارزمية الانتشار الخلفي، تتعلم شبكات الارتباط المتتابع لتقريب (لأداء) أي وظيفة بسلوك حسن مسؤول، وعند أي درجة دقة مطلوبة (Drago وRidela عام 1991[207]). فقد أثبتت أن الخطأ التربيعي التكاملي له سرعة تقارب من رتبة $O(1/n_h)$، حيث $n_h$ هو عدد عقد الطبقة المخفية.

إن خوارزمية الانتشار السريع المستعملة غالباً لتدريب مثل هذه الشبكات (Fahlman عام 1988[61]) تَستعمل المعلومات حول تغير الوزن السابق وقيمة الميل. الميل هو مجموع

المشتقات الجزئية للخطأ بالنسبة إلى الوزن المعطى عبر كل نماذج التدريب، ويعرّف كما يلي:

$$S(t) = \sum_{p=1}^{P} \frac{\partial E(p)}{\partial w} \quad (7.10)$$

وباستعمال خوارزمية الانتشار الخلفي، يعطى الميل للوزن من وحدة مخفية إلى وحدة خرج بـــ:

$$S_{jk}(t) = -\sum_{p=1}^{P} \delta_k(p) V_j(p) \quad (8.10)$$

وبالمثل، يعطى الميل للوزن من وحدة الدخل إلى وحدة مخفية بـــ:

$$S_{ij}(t) = -\sum_{p=1}^{P} \delta_j(p) x_i(p) \quad (9.10)$$

يعرَّف تغير الوزن الجديد بـــ:

$$\Delta w(t) = \frac{S(t)}{S(t-1) - S(t)} \Delta w(t-1) \quad (10.10)$$

ويمكن تعريف تغير الوزن الأولي بـــ:

$$\Delta w(0) = -\alpha S(0) \quad (11.10)$$

حيث $\alpha$ معدل التعليم.

وهكذا فإن الخطوة الأولى في خوارزمية الانتشار السريع هي ببساطة كمية تحديث من أجل الانتشار الخلفي. هناك ثلاثة حالات يجب أن تلاحَظ في تحليل سلوك هذه الخوارزمية:

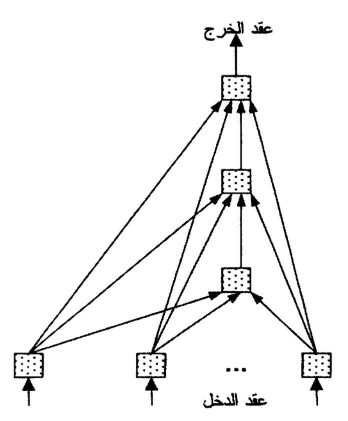
1. إذا كان الميل الحالي في نفس الاتجاه كالميل السابق ولكنه أصغر في المطال من الميل السابق، عندئذ سيكون تغير الوزن في نفس الاتجاه كما نفّذ في الخطوة السابقة.
2. إذا كان الميل الحالي في الاتجاه المعاكس للميل السابق، عندئذ سيكون تغير الوزن في الاتجاه المعاكس لتغير الوزن المنفّذ في الخطوة السابقة.
3. إذا كان الميل الحالي في نفس الاتجاه كالميل السابق ولكنه بنفس الحجم أو أكبر في المطال من الميل السابق، عندئذ سيكون تغير الوزن لانهائياً أو سيتحرك الوزن بعيداً عن القيمة الصغرى وباتجاه القيمة العظمى للخطأ. لمنع هذا المشكلة (في الحالة الثالثة) من الحدوث، يحدَّد تغير الأوزان إذا كان كبيراً جداً باستعمال عامل أزمنة الخطوة السابقة بدلاً من التغير المعطى بواسطة الصيغ لـــ $\Delta w(t)$.

## 6.10 شبكات أخرى للنمو ذاتي Other self-growing networks

سنصف في هذا المقطع شبكتين إضافيتين للنمو الذاتـــي لهما خواص مشابهة للشبكات الموصوفة فيما سبق. هاتان الشبكتان هما الشبكة البرجية والشبكة الهرمية المقترحتان من قبل Gallant عام 1990 [208].

### 1.6.10 الشبكة البرجية The tower network

تبنـــي خوارزمية الشبكة البرجية هذه الشبكةَ تزايدياً بواسطة إضافة عقد طبقة مخفية متتالية حتـــى يقلّ الخطأ إلى مستوى مقبول. توصل أول عقدة طبقة مخفية مضافة وصلاً كاملاً مع عقد الدخل ودخل الانحياز. وكذلك تتصل اتصالاً كاملاً عقد الطبقة المخفية المتتابعة مع عقد الدخل وعقدة الطبقة المخفية التـــي أسفل منها مباشرة كما هو موضح في الشكل (14.10).




**الشكل 14.10:** شبكة برجية متتالية نموذجية.

درَّبت الشبكةُ عقدةً واحدة عند لحظة ما باستعمال خوارزمية المحفظة rachet. بعد

تدريب كل عقدة، يجري تجميد n+2 وزناً (بما في ذلك دخل الانحياز) ويصبح خرج العقدة المضافة دخلاً للعقدة التالية. تعتبر عملية التدريب هذه تعديلاً لخوارزمية تعليم البيرسبترون التـــي لا يمكن أن تطبَّق إلا على شبكات بطبقة وحيدة. ولما كانت أوزان كل عقدة مضافة مثبتة عند كل مرحلة تدريب، فإن الإجراء سيكون مكافئاً لتدريب شبكة بطبقة مفردة. إن هذا النوع من الشبكات قادرٌ على تعلّم حل مسائل التصنيف غير الخطية، وقد أثبت Gallant عام 1993[22] تقارب هذه الشبكات.

### 2.6.10 الشبكة الهرمية The pyramid network

الشبكة الهرمية مشابهة تماماً للشبكة البرجية ماعدا أن كل عقدة جديدة مضافة تستقبل مداخل من كل العقد الأخرى وليس من العقدة التـــي أسفل منها مباشرة. لقد أُثبت تقارب هذه الشبكة بوجه مشابه للشبكة البرجية. يمكن أن يتوقع المرء أن الوصلات الإضافية للشبكة ستعطي زيادة في حساباتها، ولكن هذا لم يثبت بعد. سنناقش فيما يلي خوارزمية التكوين النهائية التـــي تدعى خوارزمية الانطلاق.

### 3.6.10 خوارزمية الانطلاق The upstart algorithm

اقترح Frean خوارزمية بناء هامة عام 1990[209]، حيث تضاف العقد لتصحيح الأخطاء كلما وجدت في الشبكة المستعملة. إن المداخل وتفعيلات مخارج عقد الشبكة هي عقد بعتبة خطية(ثنائية أو ثنائية القطبية). يبدأ التدريب بعقدة مفردة $u_1$ متصلة مع n عقدة دخل، وتُعدَّل الأوزان باستعمال خوارزمية المحفظة مع/بدون rachet.

تضاف عقدة جديدة إذا كانت هذه العقدة غير قادرة على تعلم واحد أو أكثر من نماذج التدريب لتصحيح الأخطاء. العقدة الإضافية ستكون عقدة $u_1^+$ أو عقدة $u_1^-$ لتعطي تقوية، موجبة أو سالبة على الترتيب، للعقدة الموجودة عندما يكون خرجها غير صحيح.

تتصل العقدة المضافة مع المداخل اتصالاً كاملاً، وتتصل مخارجها مع العقدة الأصلية $u_1$ من خلال وزن موجب كبير (للعقدة $u_1^+$) أو وزن سالب (للعقدة $u_1^-$) لتصحيح الخطأ الموافق. ليكن $T_p$ الخرج المنشود لنموذج التدريب رقم p، عندئذ تكون نماذج التدريب الموجبة للعقدة $u_j^+$ هي النماذج المعرفة بواسطة:

$u_j^p = 0$ و $T_p=1$

حيث $u_j^p$ يشير إلى خرج العقدة $u_j^+$ عندما يكون $T_p$ هو النموذج المنشود. ونماذج التدريب السالبة للعقدة $u_j^+$ هي النماذج المعرفة بواسطة $T_p = 0$. والنماذج الأخرى ستكون متجاهَلة.

وبالمثل، تعرَّف نماذج التدريب الموجبة للعقدة $u_j^-$، بواسطة:

$u_n^p = +1$ و $T_p= 0$

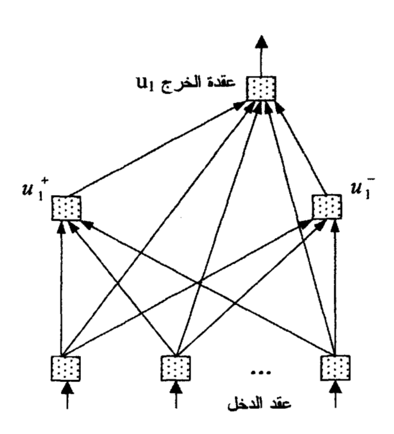
وتعرف نماذج التدريب السالبة بواسطة: $T_p=1$

تستعمل نماذج التدريب السالبة والموجبة لتصحيح الأخطاء الموجودة خلال التدريب.

يبدأ التدريب مع أول عقدة $u_1$. إذا فعلت هذه العقدة أية أخطاء "on" تضاف عقدة بنت $u_j^-$، ومن ثم تدرب هذه العقدة لتصحيح الخطأ on. إذا فعلت $u_1$ أخطاء "off" عندها تضاف عقدة بنت $u_j^+$ وتدرب هذه العقدة لتصحيح الخطأ off. بعد تدريب هذه الوحدات تجمَّد أوزانها ويوصل خرجها إلى دخل $u_1$.

إن وزن العقد $u_j^-$ هو وزن سالب كبير، تزيد قيمته على مجموع أوزان العقد الموجبة $u_j$ الأب. ووزن العقد $u_j^+$ هو وزن موجب كبير تزيد قيمته على مجموع أوزان العقد الموجبة $u_j$ الأب. تستمر هذه العملية مع العقد الإضافية الجديدة التـــي تؤدي دور الأب وتضاف حسب الحاجة عقدة التقوية: البنت السالبة أو الموجبة. تولد الوحدات الجديدة فقط إذا فعل الأب أخطاء.

من الواضح أن عدد الأخطاء ينخفض عند كل فرع للشبكة. ومن ثَم يجري إثبات التقارب باتباع مناقشات مشابهة لخوارزمية البرج، بافتراض عدم استعمال أمثلة تدريب متناقضة. الشبكة موضحة في الشكل (15.10) بعد إضافة عقدتين ابنتين، العقدة $u_1^+$ والعقدة $u_1^-$.




**الشكل 15.10**: شبكة لخوارزمية تعليم الإقلاع بعد إضافة عقدتين ابنتين

على الرغم من استعمال خرج وحيد فقط في المثال العلوي، فإن الخوارزمية يمكن أن تعمَّم لمخارج متعددة. عندما اختبرت خوارزمية الانطلاق على مسألة التكافؤ (التطابــق) بــ n-bit لمداخل ضمن مجال حتـــى 10، كانت الشبكة المكوّنة بوجه دائم بــ n وحدة بما في ذلك وحدة الخرج. في هذه التجارب، 1000 مرور عبر مجموعة التدريب كان كافياً لبناء شبكة بحجم أصغري (في حالة n = 10 كان يتطلّب 1000 مرور). نفّذت الشبكة جيداً على محاكيات مسائل أخرى بالإضافة إلى مسائل التكافؤ (التطابق).

## 7.10 تطبيقات شبكات النمو الذاتي

## Self-growing network applications

سنصف في هذه الفقرة تطبيقات عديدة لشبكات النمو الذاتـــي بأنواعها المختلفة المذكورة من قبل لتقديم حل لمسائل العالم الحقيقي. شملت هذه التطبيقات تعرّف غرض غير متغير، وحساب عدد الأسماك في مزارع السمك، وتشخيص الكبد غير العادية من صور

التصوير فوق الصوتية، وتعرّف الأحرف لتعريف أفلام أشعة X، وكشف الخلل في أسطح المنتجات المعدنية الأسطوانية الشكل، وفي تصنيف أنواع النباتات من أجل التطبيقات البيولوجية، حيث استخدمت في جميع هذه التطبيقات شبكات طاقة كولومب المخفضة.

أما شبكات الارتباط المتتابع فقد طبقت في قضايا عديدة لحل أي مسألة يمكن أن تطبق فيها الشبكات المتعددة الطبقات الأمامية التغذية بما في ذلك التطبيقات المذكورة في الفصل السابع وخاصة المسائل الصعبة في تعرف الأشكال، ومسائل التنبؤ المالي وسنعرض الآن بعضاً من هذه التطبيقات.

### 1.7.10 تعرف غرض غير متغير Invariant object recognition

يعتبر تطبيق الرؤية الحاسوبية إحدى التطبيقات التـــي نحتاج إليها فيها تعرف أغراض عديدة كلما تحركت على طول السلسلة الناقلة (لهذه الأغراض، وخاصة في خطوط الإنتاج الصناعية). وهذا يتطلب أن تكون هذه الأغراض مميزة عند محلات عديدة في حقل الرؤية وبمقاييس مختلفة وباتجاهات مختلفة أيضاً.

استعمل Wei Li وNasrabadi عام 1990 [210] ثلاث شبكات طاقة كولومب المخفضة متتابعة لتعلم مسألة تعرف غرض غير متغير. لتنفيذ اللاتغيرية عولجت سلفاً صورة ذات قياس 256×256 عنصر صورة بشدات مستويات رمادية ضمن المجال (0-255) وطبقت إلى شعاع الملامح ذي البعد 48. بعد ذلك عولجت الملامح من قبل الشبكة لتعرُّف الأغراض.

تألفت المعالجة المسبقة لمعطيات الصورة من وضع قيم عتبات للصورة لتحويلها إلى صورة رقمية ثنائية باستعمال طرق المدرج التكراري للشدات، وحساب المركز المتوسط للغرض المحول إلى الشكل الثنائي، والتحويل من الإحداثيات المتعامدة إلى الإحداثيات القطبية، وبعدئذ معايرة الصورة. قُسِّم تمثيل الغرض بعدئذ إلى 12 مقطعاً زاوياً (30/360) حول المبدأ واستنبطت رموز الواصفات الأربعة من كل مقطع.

الأجزاء الأربعة للمعلومات هي:

1. عدد نقاط حدود الغرض في المقطع مقسوماً على العدد الكلي لنقاط الحدود
2. المساحة المعيارية للغرض في المقطع (العدد الكلي لعناصر الصورة السوداء)

3. المسافة العظمى لحدود الغرض عن نقطة المركز المتوسطة لكل مقطع
4. المسافة الصغرى لحدود الغرض عن المبدأ.

توفر هذه الحسابات للمعالجة القبليّة اللاتغيرية لانزياح الغرض، والمقياسَ، والتوجيهَ. طبقت مسألة تعرف الأشكال اللاتغيرية على حركة مطرقة، حيث أخذت ثلاثة صور من أماكن مختلفة. لإنجاز تعميم أفضل في التعليم، درِّبت ثلاث شبكات طاقة كولومب المخفضة متتالية باستعمال عدة أدوات مثل مطرقة، وكماشة، ومفتاح إنكليزي، وغير ذلك من الأدوات.

استُعملت مجموعة تدريب مؤلفة من 48 ملمحاً. خلال التدريب، إذا أخفقت أول شبكة في تعرُّف بعض الملامح تعرفاً صحيحاً بعد أدوار متعددة يوقف التعليم وتربط شبكة جديدة على التتالي مع الشبكة الأولى لتدرَّب على تعرف الملامح غير المصنفة. وهكذا تتكرر العملية حتـــى يصنف تتابع الشبكات بوجه صحيح كل الملامح، يلزم لهذا التطبيق ثلاث شبكات فقط.

درِّبت الشبكة الأولى وعلِّمت على ملامح الغرض الكبيرة (ليست الدقيقة)، أما الشبكتان الباقيتان فقد عُلِّمتا على الملامح الدقيقة والناعمة. رُكِّبت الشبكة الأولى من 11 عقدة طبقة داخلية وعلّمت على تصنيف 34 ملمحاً، وركبت الشبكة الثانية من 14 عقدة طبقة داخلية وعلّمت على تصنيف 13 ملمحاً، أما الشبكة الثالثة فقد ركبت من عقدة طبقة داخلية واحدة وعلّمت على تمييز الملمح الأخير المتبقي. خلال طور تعرّف الأغراض، استُعمل 50 منظراً مختلفاً للأغراض لاختبار الشبكة باستعمال آلة تصوير ذات اختيار عشوائي للوضعيات واتجاهات الأغراض. نفّذت الشبكة تصنيفاً بدقة 100% في هذا التطبيق، وأثبتت خوارزمية طاقة كولومب المخفضة سرعةً في التدريب والتعرف مقارنة مع بين الشبكات الأخرى.

### 2.7.10 حساب عدد الأسماك Fish counting

مزارع السمك في نمو متزايد في أرجاء العالم. حيث تعتبر مصدراً هاماً للبروتين العالي والطعام قليل الدسم. المشكلة الكبيرة التـــي تواجه مزارعي السمك هي تعقب أثر الأسماك، وحساب كثافة الـــ minnows في عينة مأخوذة من محلات خاصة. على الرغم من تنفيذ

طرق عديدة آلية، فإن طريقة التقدير البشري تعتبر الأكثر دقة حتـــى اليوم، مع أن هذه الطريقة تعطي 10% معدل خطأ.

إن حلول شبكات طاقة كولومب المخفضة تبدي بعض الآمال في دقة أفضل (Collins عام 1992[28]). بنـــي حل الشبكة العصبونية على حساب عدد الـــ minnows التـــي تظهر في الصورة المأخوذة لعينة مسكوبة في حوض قليل العمق. قدر العدد الكلي للـــ minnows بجمع تجمعات الـــ minnows التـــي تظهر في مناطق مختلفة من المساحة. كل تجمع يمكن أن يحوي من (0.5) minnows.

في البداية عولجت الصورة سلفاً بتحويلها إلى الشكل الرقمي الثنائي وفعل لها تعرية لإزالة الفقاعات العضوية الصغيرة. فقاعات عناصر الصورة المتصلة استعملت بعدئذ لحساب ملامح الأغراض مثل؛ المساحة، قياس الجوار، التشتت، قياس القطر، عدد الرؤوس المدببة، ...الخ. عملت هذه الملامح كدخل لشبكة طاقة كولومب المخفضة التـــي ستعطي عدد minnows في التجمع.

لتدريب الشبكة، عملت الملامح المحسوبة للتجمع كدخل للشبكة مع العدد الفعلي المراقب بواسطة المُشغِّل؛ شاشة فيديو متصلة مع إطار نزع (أخذ) يعمل كجزء من نظام الرؤية، عندما يكتشف التجمع يضاء النظام بشدة ويطلب من المشغل العدد الفعلي في التجمع. تُستعمل هذه العينة لتدريب الشبكة.

### 3.7.10 تشخيص الكبد Liver diagnosis

يمكن أن تصنف الكبد البشرية بكونها عادية أو غير عادية من خلال تحليل الصور فوق الصوتية للكبد عند الإنسان. استعملت ملامح إحصائية وهندسية لتمييز صور الكبد، مثل: الشدة المتوسطة، والشدة المعتدلة، وتردد المستوى الرمادي الأعظمي، وقيم الشدة الدنيا، وقيم الشدة العظمى، وتباين الشدة، والانحراف القياسي، والنعومة، ونقاط النسب المئوية للشدة من 10% حتـــى 90 %، وعامل الاحتمال الأعظمي، وعامل الأنتروبي، وعامل التفاضل العكسي، وعوامل الارتباط، والتـــدرج المتوسط، والانحراف القياســـي للتـــدرج، ... الخ (Collins عام 1992[28]).

فحصت طرق تصنيف عديدة وقورنت نتائجها بما في ذلك تجمعات الجوار الأقرب، وتحليل التمييز الخطي، والشبكات العصبونية المتعددة الطبقات الأمامية التغذية المدربة بالانتشار الخلفي، وشبكات الانتشار المتعاكس (ستدرس لاحقاً)، وشبكات طاقة كولومب المخفضة. وكما نشر لم تستطع أية تقنية أن تضاهي أو تنافس الدقة 90% المعطاة بواسطة شبكة طاقة كولومب المخفضة.

### 4.7.10 تعرف أحرف أفلام أشعةX

تُحدَّد نوعية أفلام أشعة X الخاصة بمريض باستعمال أعداد التعريف ID المسجلة على قاعدة الفيلم، وتُستعمل هذه الأعداد لتخزين ونقل الأفلام في المشافي. وهذا يتطلب وجود طرق سريعة وموثوقة لتعرّف الأحرف آلياً وذلك لتسريع وظائف التخزين والنقل.

استُعملت شبكة طاقة كولومب لإنجاز عمليـــة التـــعرف بمعـــدل دقـــة تميـــيز 99.4% (Hasegawa عام 1993[211]). تضم الأحرف المستعملة لتعريف المرضى الأحرف T وM وS وعشر خانات عشرية من الصفر وحتـــى التسعة. سُجلّت أيضاً أسماء المرضى ولكن لم تستعمل في عملية التعريف ID.

تشمل عملية المعالجة القبليّة لصورة التعريف ID تكوين المقاطع، والدوران، والتحويل إلى الشكل الرقمي الثنائي.

يتألف الحرف المعالج سلفاً من 12×15 عنصر صورة استعملت كدخل للشبكة. وهكذا فإن للشبكة 180 مدخلاً بقيمة ثنائية و13 مخرجاً (عشر أرقام + ثلاثة أحرف). حُدِّد عدد وحدات الطبقة المخفية آلياً خلال طور التدريب.

تـــدعـــى خـــطـــة التـــعليـــم المـــعـــدلة المستعمـــلـــة للمســـألـــة HSII؛ (Hyper–spherical Surface Interactive Interconnections).

تَستعمل هذه الطريقة نوعاً من الانتشار الخلفي لتهذيب الأوزان الواصلة إلى مداخل الطبقة المخفية. وقد وُجد أنها أقوى من خوارزمية تعليم طاقة كولومب المخفضة الأساسية.

بوجه عام، تولّد طريقة HSII عدد عقد مخفية أقل من خوارزمية طاقة كولومب المخفضة الأساسية خلال عملية التعليم، وتحقق إنجازاً أقوى.

مثلاً، أعطت هذه الطريقة 31 عقدة مخفية مقارنة مع 55 عقدة مخفية أعطيت من قبل طاقة كولومب المخفضة الأساسية، وكذلك زمن التعليم كان أقصر لطريقة HSII.

دربت الشبكة على مجموعة مؤلفة من 2250 حرف أخذت من 90 صفيحة أشعة X. تحقق التقارب بعد تنفيذ 710 عملية تكرار.

## 5.7.10 تعرف الأشكال باستعمال شبكة الارتباط المتتابع

تعتبر مسألة ثنائي اللولب واحدة من أصعب علامات اختبار مقدرة إنجاز الشبكات. إن مداخل الشبكة لهذه المسألة هي نقاط بإحداثيات x-y بقيم مستمرة تصف لولبان مدموجان في فراغ ثنائي البعد.

الشبكة لها عقدة خرج وحيدة تعطي +1 للنقاط المرافقة للولب الأول وتعطي -1 للنقاط المرافقة للولب الثاني. من الواضح أنه يمكن أن تعلم عقد الطبقة المخفية اللازمة في أي شبكة لحل هذه المسألة ذات الفصل غير الخطي.

درست هذه المسألة باستعمال شبكة ارتباط متتالٍ من قبل Fahlmann وLeibiere عام 1990[205]. نفذت المسألة 100 مرة باستعمال توابعَ تفعيل sigmoid لكل العقد وحوض وحدات الطبقة المخفية الثمانية. كل التجارب كانت ناجحة وقد تطلبت 1700 تدريب وسطياً.

بلغ عدد العقد المخفية المركبة للشبكة النهائية من 12 حتـــى 19. وكانت أزمنة التدريب أقل بعامل عشرة مما هي عليه في خوارزمية الانتشار الخلفي، على حين كان تركيب الشبكة بنفس التعقيد (15 عقدة طبقة مخفية في الانتشار الخلفي).

نُفِّذت تجارب ثنائية اللولب أخرى باستخدام شبكات الانتشار الخلفي التقليدية ولكنها تطلبت حوالي 150000-200000 تدريب! (تقرير لم ينشر لشركة Alexis Wieland of MITRE) [28]. ونفذت تجارب على مسائل أخرى أعطت نتائج درامية مشابهة.

# شبكة النيوكونيترون
# Neocognitron

سنناقش في هذا الفصل نوعاً غير عادي من بنـى الشبكات العصبونية الصنعية التـي صممت أساساً لتنفيذ مهام معالجة الصور، كمسألة تعرّف الأشياء بوجه مستقل عن المكان والأحرف المكتوبة يدوياً. وقد جرى بعد اختراع هذه الشبكة تهذيب بنيتها عبر السنين، وتطويرها لتتوسع مقدراتها وليتحسن إنجازها لكي تقوم بنمذجة قريبة من نظام الرؤية البشري.

إن الخواص التـي جعلت من هذه الشبكة فريدة هي نظام التوصيل الاختياري بين الطبقات المرتبة تتابعياً، حيث يجري كشف معالم المستوى المنخفض البسيطة للشيء المطلوب تعرفه في الطبقات الأولى، ثم تركب هذه الملامح البسيطة المكتشفة لتكوين الشيء بشكل كامل في الطبقات المتعاقبة من خلال توسيع منطقة الحقل المستقبل للدخل.

جرى اعتماد أفكار تشكيل الوصلات في شبكات أخرى عديدة لتنفيذ مقدرات معالجة رؤية خاصة.

يتضمن الفصل مقطعاً تمهيدياً ووصفاً للبنية وخوارزميات التدريب وطريقة عمل النيوكونيترون، ومن ثم شرح النسخة المعززة للشبكة المتضمنة وصلات تغذية عكسية. وفي المقطع الأخير سنقدم بعض التطبيقات النموذجية لهذه الشبكة.

## 1.11 تمهيد

أحد أهم الأجهزة المستعملة في تعرف الأشكال وتمييزها هو نظام الرؤية عند الإنسان. وهذا النظام جدير بالملاحظة؛ إذ يستطيع الإنسان بواسطة هذه الشبكة المعقدة جداً من الحساسات وعصبونات معالجة الإشارة تعلم الأشياء وتمييزها من بين خلفيات متنوعة مستقلة

عن المكان النسبـــي والحجم والاتجاه، حتـــى إنه يمكن تعرف الأشياء الظاهرة جزئياً من رؤية جزء بسيط من هذه الأشياء.

إنه تحدٍ حقيقي للباحثين في حقل الرؤية الحاسوبية لمقدرتهم على إنجاز مستويات مشابهة بواسطة نماذج أنظمة الرؤية الصنعية.

مازالت إنجازات الباحثين في هذا المضمار متواضعة مقارنةً مع جهاز الرؤية عند الإنسان، مع أن بعض النتائج الواعدة بدأت تلوح في الأفق في مطلع التسعينيات.

لقد أظهرت إحدى بنـــى الشبكات العصبونية الصنعية المسماة نيوكونيترون مقدرة في تعرف الأشكال غير المتغيرة والمحددة؛ وهذا يعني أن تمييز الأشياء مستقلة عن أماكنها في الصورة ومستقلة عن التشوهات الشكلية التـــي قد تعتري الأشكال أو حتـــى عند الظهور الجزئي لهذه الأشكال.

اقتُرحت شبكة النيوكونيترون من قبل الباحث اليابانـــي **Fukushima** عام 1982[8] وجرت دراستها والاستمرار بتعديلها من قبل مخترعها حتـــى مطلع 1991[185] [184] [212].

طُوِّرت هذه الشبكة عن نموذج عصبونـــي قديم متكيف ذاتياً بطبقات متعددة يسمى كونيترون (cognitron) (**Fukushima** عام 1975[213]). في البداية اقتُرح هذا النظام القديم نموذجاً لتعرف الأشكال البصرية في الدماغ. وهو عبارة عن شبكة أمامية التغذية مدربة بدون معلم (تعليم بدون معلم)، وكان قادراً على تعلم مهام تمييز الأشكال المعقدة وتعرُّفها.

إن شبكة الكونيترون مثل الكثير من بنـــى الأنظمة البصرية الأخرى، حساسة للإزاحة والمقياس وتشوهات أخرى في الصورة. مثلاً، إذا دُرِّبت هذه الشبكة على تمييز شيء، وليكن شخصاً بلحية وقبعة، في مكان ما من الصورة، فإنها لن تستطع تعرفه ثانية عند ظهوره في مكان آخر، أو حتـــى عندما يعتري هذا الشيء بعض التشويه؛ كخلع القبعة أو حلق اللحية.

طُوِّرت شبكة النيوكونيترون على التالي عبر السنين للتغلب على هذه الصعوبات، حيث أصبحت قادرة على تعلم الإزاحة ومهام تمييز الأشكال غير المتغيرة والمشوّهة أيضاً. وقد استُخدم التعليم بمعلم وبدون معلم في تدريب هذه الشبكات.

صممت الشبكة لتعرّف أحرف الكتابة وخاصة الأرقام العربية 0, 1, 2, ..., 8, 9 (سيشرح فيما بعد). كان الهدف من الشبكة جعلها تستجيب دون أن تتأثر بالتغيرات في

المكان والشكل وخط الطباعة الذي تكتب فيه الأرقام والأحرف.

تتألف بنية النيوكونيترون من طبقات عديدة من الوحدات. رُتِّبت الوحدات ضمن كل طبقة في عدد من المصفوفات المربعة. تَستقبل الوحدة ضمن مصفوفة من مصفوفات الطبقة الواحدة إشارات من عدد محدود جداً من الوحدات في الطبقة السابقة، وبالمثل ترسل الإشارات إلى بضع وحدات فقط في الطبقة التالية. ورُتِّبت وحدات الدخل في مصفوفة مربعة وحيدة تضم 19 × 19 وحدة، وهي عصبونات مستقبلة ضوئية؛ (photoreceptors neurons) أي خلايا مستقبلة حساسة للضوء. قد تكون مخارج الوحدات ثنائية أو بقيمة حقيقية (مستوى رمادي) عند استجابتها لصورة الدخل. حجم المصفوفة قد يكون عدداً صغيراً من الوحدات (8 × 8) وقد يكون كبيراً حتـــى 128 × 128 وحدة. الاختيار العام لحجم المصفوفات يقع في المجال 16 × 16 إلى 32 × 32 ومخارج الوحدات تكون ثنائية القيمة.

في مسألة تعرف الأرقام العربية التـــي نعالجها، تتألف الطبقة الأولى التـــي تلي طبقة الدخل من 12 مصفوفة مربعة، يتألف كل من هذه المصفوفات من 19 × 19 وحدة. بوجه عام، يتناقص حجم المصفوفات كلما تقدمنا من طبقة الدخل باتجاه طبقة خرج الشبكة، وسنصف البنية بالتفصيل فيما يلي.

رتِّبت الطبقات أزواجاً، كل زوج يضم طبقتين، الطبقة S ("S" للخلايا البسيطة، كما وجدت في القشرة البصرية الأولية) متبوعة بالطبقة C ("C" للخلايا المعقدة، كما وجدت في القشرة البصرية). وزِّعت الوصلات من الدخل إلى الطبقة S الأولى اختيارياً بأسلوب خاص، وكذلك الوصلات من الطبقات S إلى الطبقات C، وسيوضَّح ذلك لاحقاً. سيختلف العدد الفعلي للطبقات في الشبكة وفقاً للمسألة المعالجة.

عموماً، تتطلّب النماذجُ المعقدة المطلوب تعرُّفها وتمييزها شبكاتٍ بعدد أكثر من الطبقات، ويتطلّب تعرف عدد أكبر من الأشياء عدداً أكبر من الطبَقات المركبة من مصفوفات عامرة بالوحدات الغزيرة (الكثيفة). درِّبت مصفوفات الطبقة S لتستجيب لنموذج خاص (ملمح أو سمة) أو مجموعة من النماذج (مصفوفات استنباط المعالم)، ومن ثم تقوم مصفوفات الطبقة C (المحشورة بين الطبقات S) بكشف الأخطاء (الإزاحات) المكانية في المعالم.

نتيجة التناوب في الطبقات S وC يجري استنباط المعالم من قبل الطبقات S، ومن ثم

تـــركيب هذه المعالم البسيطة مع ملاحظة إزاحاتها المكانية (ضم النتائج من مصفوفات الطبقة S)، ويخفّف أو يرقق لحظياً عدد الوحدات في كل مصفوفة.

سنشرح الحاجة إلى نسخ متعددة من المصفوفات في كل طبقة عند شرح تدريب الشبكة.

ببساطة نلاحظ أن كل مصفوفة (ضمن الطبقة الواحدة) درِّبت لتستجيب لنموذج مختلف من العلامات أو السمات الفارقة (معالم الدخل الأصلي)، حيث تبحث كل وحدة في مصفوفة خاصة عن تلك السمة في جزء صغير من الطبقة السابقة.

يتقدم التدريب طبقة بعد طبقة، حيث تدرب الأوزان الواصلة من وحدات الدخل إلى الطبقة الأولى ومن ثم تجمّد. بعد ذلك تعدّل الأوزان التالية القابلة للتدريب وهلم جراً. تثبت الأوزان بين الطبقات، وكذلك نماذج الوصل، عندما يتم تصميم الشبكة نهائياً.

## 2.11 بنية النيوكونيترون

تتألف بنية شبكة النيوكونيترون من تسع طبقات، بعد طبقة الدخل هناك أربعة أزواج من الطبقات. تتألف أول طبقة، من كل زوج، من المصفوفات S، والطبقة الثانية من المصفوفات C. سنرمز للطبقات حسب ترتيبها بما يلي:

دخل (U), S1, $C_1$, $S_2$, $C_2$, $S_3$, $C_3$, $S_4$, $C_4$ وهي تمثل طبقة الخرج. رتبت الوحدات في كل طبقة ضمن مصفوفات مربعة عديدة (أو خلايا) وفقاً للجدول التالي (في مسألة تعرف الأرقام العربية):

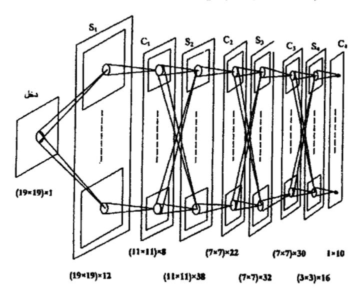
| الطبقة | عدد المصفوفات ضمن الطبقة | حجم كل مصفوفة |
|---|---|---|
| دخل (U) | 1 | 19 ×19 |
| $S_1$ | 12 | 19 × 19 |
| $C_1$ | 8 | 11 × 11 |
| $S_2$ | 38 | 11 × 11 |
| $C_2$ | 22 | 7 × 7 |
| $S_3$ | 32 | 7 × 7 |
| $C_3$ | 30 | 7 × 7 |
| $S_4$ | 16 | 3 × 3 |
| $C_4$ | 10 | 1 × 1 |

ويوضح الشكل (1.11) بنية هذه الشبكة.

نلاحظ من البنية المفصلة لأول ثلاث طبقات من الشبكة المبينة في الشكل (2.11)، أن الطبقة $S_1$ مؤلفة من مصفوفات مربعة من الوحدات، وتؤدي هذه المصفوفات دوراً هاماً في كشف معالم المستوى المنخفض في الصورة.

دربت كل مصفوفة لتستجيب لمعلم نموذج مستوى منخفض منفصل ومختلف عن معلم المصفوفات الأخرى. لذا سيكون للشبكة مصفوفات أكثر في الطبقة $S_1$ من عدد المعالم الأولية المطلوب تمييزها في طبقة الدخل.

تكون كل الوحدات ضمن مصفوفة واحدة من مصفوفات Sمتجانسة، أي تستجيب لنفس المعلم. ينجز هذا بوصل كل خلية في المصفوفة بمجموعة صغيرة من الوحدات في الطبقة السابقة، حيث لكل مجموعة التوزيع المكانـــي نفسه من العناصر في الطبقة السابقة، ولكنها تأتـــي من منطقة مختلفة (إزاحة متوازية) في الطبقة.

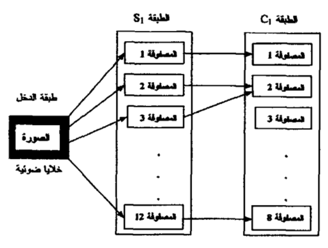



**الشكل1.11**: بنية النيوكونيترون، تشير الرموز العلوية إلى أسماء الطبقات، والأعداد السفلية إلى حجم كل طبقة من الوحدات (عدد المصفوفات في كل طبقة × حجم المصفوفة (عدد الأسطر × عدد الأعمدة) وحدة) .[212]

سنرمز للمصفوفة (أو الخلية) ضمن الطبقة الواحدة بدليل علوي؛ أي ستكون المصفوفة الأولى في الطبقة $S_1$ هي $S_1^1$، والمصفوفة الثانية في الطبقة $S_1$ هي $S_1^2$، وهكذا. ويشار إلى الوحدات ضمن مصفوفة خاصة بواسطة أدلة سفلية.

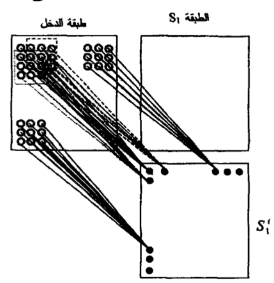
تَستقبل كل وحدة في واحدة من المصفوفات إشارات من مجموعة صغيرة من الوحدات

في الطبقة السابقة. مثلاً، الوحدة في المصفوفة الأولى من الطبقة $S_1$ هي $S^1_{1i,j}$ (مع ملاحظة كون الوحدة من الطبقة C أو من الطبقة S) ستستقبل إشارات من وحدات مخصصة في مصفوفة أو أكثر من الطبقة السابقة.




الشكل2.11: الطبقات الأولى S وC مع مصفوفاتها الموافقة

لجعل هذه الفكرة ملموسة على الواقع، سنناقش في البداية الوحدات في طبقة الدخل المرتبة في مصفوفة مربعة بحجم 19 × 19 والوحدات في الطبقة $S_1$. تتألف الطبقة $S_1$ من 12 مصفوفة مربعة هي $S_1^1$، $S_1^2$، ... $S_1^{12}$ كل منها بحجم 19 × 19 وحدة. تستقبل وحدة في مصفوفة واحدة من مصفوفات $S_1$إشارات من مصفوفة بحجم 3 × 3 وحدة (هذه هي المنطقة الصغيرة المذكورة سابقاً) في طبقة الدخل، كما هو موضح في الشكل (3.11)




الشكل 3.11: وصلات الدخل إلى المصفوفة $S_1^i$ من مصفوفات الطبقة $S_1$

تظهر في الشكل خمس مجموعات من الوصلات فقط لتوضيح الحقول المستقبلة لوحدات مختلفة في المصفوفة $S_1^i$من مصفوفات $S_1$. لإتمام النموذج التوصيلي في الشكل (3.11)، ستكون وصلات الوحدة الثانية في الزاوية اليسرى العليا في المصفوفة $S_1^i$مـــن مجموعة مربعة 3×3 من الوحدات في طبقة الدخل المزاحة عموداً واحداً إلى اليمين عن الزاوية اليسرى العليا، كما هو مبين في الشكل (3.11). تستقبل الوحدة الثالثة من الزاوية اليسرى العليا من المصفوفة $S_1^i$إشارات من مجموعة 3 × 3 مربعة مزاحة عمودين إلى اليمين عن الوحدة في الزاوية العليا اليسرى، وهكذا حتـــى الوصول إلى المجموعة 3×3 في الطرف العلوي الأيمن من طبقة الدخل. تستقبل وحدة المصفوفة $S_1^i$ أسفل وحدة الزاوية العليا اليسرى في السطر الثانـــي مداخلها من المجموعة 3×3 اليسرى العليا من الوحدات في طبقة الدخل المزاحة للأسفل سطراً واحداً، كما هو مبين في الشكل (3.11) وهكذا. يمكن أن تتصل المجموعات في أقصى اليمين من طبقة الدخل مع الوحدات في أقصى اليسار، أي يمكن أن تكون طبقة الدخل ملفوفة وذلك بوصل أعمدة (أسطر) الوحدات في أقصى اليمين (القاعدة) بالأعمدة (الأسطر) في أقصى اليسار، وهذا سيعطي دخلاً مستمراً لكل النقاط في الصورة. **كل المصفوفات في $S_1$ لها نفس نماذج التوصيل كما وصف من قبل، ولكنها تستجيب لمعالم مستوى منخفض مختلفة.**

النماذج التوصيلية من وحدات الطبقة C إلى وحدات الطبقة S مشابهة لنظام التوصيل بين الدخل والطبقة $S_1$ ولكن الحقول المستقبلة تصبح أكبر (أوسع) تتابعياً بحيث تتعلم وحدات الطبقة S التـــي هي أعمق أو أبعد استجابة إلى معالم كلية أكبر في الدخل.

مثلاً، الوحدة $S_{1i,j}^1$ تستقبل إشارات من وحدات الدخل التسع $U_{i-1,j-1}$، $U_{i-1,j}$، $U_{i-1,j+1}$، $U_{i,j-1}$، $U_{i,j}$، $U_{i,j+1}$، $U_{i+1,j-1}$، $U_{i+1,j}$، $U_{i+1,j+1}$. وكذلك تستقبل الوحدة $S_{1i,j}^2$ إشارات تماماً من نفس وحدات الدخل، وهكذا من أجل كل مصفوفة من 12 مصفوفة في الطبقة $S_1$ (حالة خاصة: إذا كان الدليل i أو j يساوي 1 أو 19 فإن وحدة $S_1$ تستقبل إشارات من أربع وحدات دخل فقط، عوضاً عن تسع، في الحالتين الخاصتين الموضحتين في الشكل (3.11)، إذا كان $i = 19, j = 19$ فإن الوحدات الأربع ستكون

، $U_{18,19}, U_{18,18}, U_{19,19}, U_{19,18}$ وإذا كان $i=1, j=19$ فإن الوحدات الأربعة هي ($U_{1,19}, U_{1,18}, U_{2,19}, U_{2,18}$). وبوجه عام، تستقبل وحدة في مصفوفة الطبقة S إشارات من وحدات مخصصة في كل مصفوفات الطبقة C السابقة.

تدعى الطبقة الثانية من كل زوج من الطبقات بالطبقة C. تعمل الطبقات C لترقيق أو تخفيف عدد الوحدات في كل مصفوفة (باستقبال دخل من حقل رؤية أعرض نوعاً ما). تستقبل المصفوفة في الطبقة C دخلاً من مصفوفة واحدة أو من مصفوفتين أو من ثلاث مصفوفات في الطبقة S السابقة، وعندما تستقبل المصفوفة إشارات من أكثر من مصفوفة S واحدة، فإن المصفوفة C تضم النماذج المتشابهة (ذات المعالم الواحدة) من الطبقة S. تتألف الطبقة $C_1$ من 8 مصفوفات مربعة حجم كل منها 11×11 وحدة.

تستقبل المصفوفة $C_1^1$ إشارات من حقل بحجم 5×5 وحدة في المصفوفة $S_1^1$. المصفوفة $C_1^2$ لا تعمل فقط على تكثيف الإشارات من منطقة الوحدات S، ولكنها أيضاً تضم الإشارات الموافقة للنماذج المتشابهة التـــي دربت عليها المصفوفة $S_1^2$ أو المصفوفة $S_1^3$. وهكذا، تستقبل الوحدة $C_1^2$ إشارات من منطقة في المصفوفة $S_1^2$ ومن نفس المنطقة في المصفوفة $S_1^3$. فيما يلي سنعتبر أولاً نموذج الوصلات بين المصفوفات عند مستويات مختلفة، ومن ثم سنصف حقل الرؤية من أجل الوحدات ضمن كل مصفوفة.

تستقبل كل مصفوفة $S_2$ إشارات من كل مصفوفات $C_1$؛ وهذا يعنـــي أن كل وحدة من مصفوفات $S_2$ تستقبل إشارات من نفس الجزء من كل مصفوفة من مصفوفات $C_1$ في الطبقة السابقة. وبالمثل، تستقبل كل مصفوفة $S_3$ إشارات من كل المصفوفات $C_2$، وتستقبل كل مصفوفة $S_4$ إشارات من كل مصفوفات $C_3$. على أية حال، وكما ذكر من قبل، تستقبل المصفوفات في الطبقة $C_1$ إشارات من واحدة فقط أو غالباً بضع، من مصفوفات $S_1$ التـــي تقع في نفس المستوى.

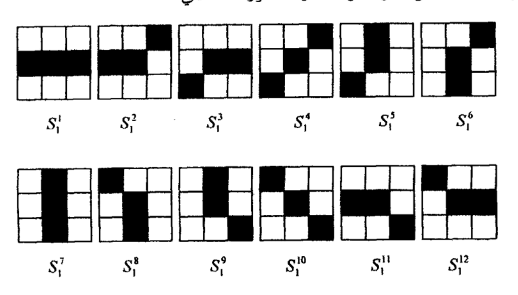
وعادةً يكون نموذج الوصل كما يلي:

**الوصلات من $S_1$ إلى $C_1$:**

$$
\begin{array}{llll}
S_1^1 & \to C_1^1 & S_1^7 & \to C_1^5 \\
S_1^2, S_1^3 & \to C_1^2 & S_1^8, S_1^9 & \to C_1^6 \\
S_1^4 & \to C_1^3 & S_1^{10} & \to C_1^7 \\
S_1^5, S_1^6 & \to C_1^4 & S_1^{11}, S_1^{12} & \to C_1^8
\end{array}
$$

يصبح الباعث لنماذج الوصل هذه أوضح إذا نظرنا إلى النماذج المستعملة لتدريب الأوزان من طبقة الدخل إلى الطبقة $S_1$. دربت المصفوفة $S_1^1$ لتستجيب لقطعة مستقيمة أفقية صغيرة كما هو موضح في الشكل (4.11). أما المصفوفتان $S_1^2$ و $S_1^3$ فقد دربتا لتستجيبا للقطع المستقيمة عند زاوية 22 درجة تقريباً عن الخط الأفقي.

تعمل المصفوفة $C_1^2$ على ضم الإشارات من هاتين المصفوفتين. وبطريقة مشابهة، تستجيب $S_1^5$ و $S_1^6$ لأشكال مختلفة من القطع بين القطرية والعمودية، ومن ثم تضم إشاراقما إلى مصفوفة وحيدة في $C_1$. كوّنت نماذج الوصل من المصفوفات $S_2$ إلى المصفوفات $C_2$ بنفس الطريقة والأفكار المذكورة كما يلي:




**الشكل 4.11**: نماذج بسيطة لتدريب الطبقة S1 لشبكة النيوكونتيرون [212]

**الوصلات من $S_2$ إلى $C_2$:**

| | | |
|---|---|---|
| $S_2^1, S_2^2, S_2^3 \rightarrow C_2^1$ | $S_2^{13}, S_2^{14} \rightarrow C_2^8$ | $S_2^{26} \rightarrow C_2^{15}$ |
| $S_2^2, S_2^3, S_2^4 \rightarrow C_2^2$ | $S_2^{15}, S_2^{16} \rightarrow C_2^9$ | $S_2^{27}, S_2^{28}, S_2^{29} \rightarrow C_2^{16}$ |
| $S_2^5 \rightarrow C_2^3$ | $S_2^{17}, S_2^{18} \rightarrow {}_2^{10}$ | $S_2^{30}, S_2^{31} \rightarrow C_2^{17}$ |
| $S_2^6, S_2^7, S_2^8 \rightarrow C_2^4$ | $S_2^{19} \rightarrow C_2^{11}$ | $S_2^{32} \rightarrow C_2^{18}$ |
| $S_2^7, S_2^8, S_2^9 \rightarrow C_2^5$ | $S_2^{20}, S_2^{21} \rightarrow C_2^{12}$ | $S_2^{33} \rightarrow C_2^{19}$ |
| $S_2^{10} \rightarrow C_2^6$ | $S_2^{22}, S_2^{23}, S_2^{24} \rightarrow C_2^{13}$ | $S_2^{34} \rightarrow C_2^{20}$ |
| $S_2^{11}, S_2^{12} \rightarrow C_2^7$ | $S_2^{25} \rightarrow C_2^{14}$ | $S_2^{35}, S_2^{36} \rightarrow C_2^{21}$ |
| | | $S_2^{37}, S_2^{38} \rightarrow C_2^{22}$ |

يحدث تركيب قليل جداً في الذهاب من الطبقة S 3 إلى الطبقة C3، حيث تضم الإشارات من المصفوفات $S_3^{23}$ و $S_3^{24}$ في المصفوفة $C_3^{23}$. وتضم الإشارات من المصفوفات $S_3^{30}$ و $S_3^{31}$ في المصفوفة $C_3^{29}$. وتستقبل كل مصفوفات C3 الأخرى إشارات من مصفوفة واحدة فقط في S3.

**الوصلات من S3 إلى C3:**

| | | |
|---|---|---|
| $S_3^1 \rightarrow C_3^1$ | $S_3^{11} \rightarrow C_3^{11}$ | $S_3^{21} \rightarrow C_3^{21}$ |
| $S_3^2 \rightarrow C_3^2$ | $S_3^{12} \rightarrow C_3^{12}$ | $S_3^{22} \rightarrow C_3^{22}$ |
| $S_3^3 \rightarrow C_3^3$ | $S_3^{13} \rightarrow C_3^{13}$ | $S_3^{23}, S_3^{24} \rightarrow C_3^{23}$ |
| $S_3^4 \rightarrow C_3^4$ | $S_3^{14} \rightarrow C_3^{14}$ | $S_3^{25} \rightarrow C_3^{24}$ |
| $S_3^5 \rightarrow C_3^5$ | $S_3^{15} \rightarrow C_3^{15}$ | $S_3^{26} \rightarrow C_3^{25}$ |
| $S_3^6 \rightarrow C_3^6$ | $S_3^{16} \rightarrow C_3^{16}$ | $S_3^{27} \rightarrow C_3^{26}$ |
| $S_3^7 \rightarrow C_3^7$ | $S_3^{17} \rightarrow C_3^{17}$ | $S_3^{28} \rightarrow C_3^{27}$ |
| $S_3^8 \rightarrow C_3^8$ | $S_3^{18} \rightarrow C_3^{18}$ | $S_3^{29} \rightarrow C_3^{28}$ |
| $S_3^9 \rightarrow C_3^9$ | $S_3^{19} \rightarrow C_3^{19}$ | $S_3^{30}, S_3^{31} \rightarrow C_3^{29}$ |
| $S_3^{10} \rightarrow C_3^{10}$ | $S_3^{20} \rightarrow C_3^{20}$ | $S_3^{32} \rightarrow C_3^{30}$ |

تتألف المصفوفة C4 من عشر مصفوفات كل منها بوحدة مفردة، وحدة واحدة لكل خانة من الخانات الرقمية العشرة التـي صممت الشبكة لتمييزها. تركب الإشارات من المصفوفات S4 لتشكل الاستجابة النهائية للشبكة. نماذج التوصيل من مصفوفات S4 إلى

مصفوفات $C_4$ كما يلي:

**الوصلات من $S_4$ إلى $C_4$:**

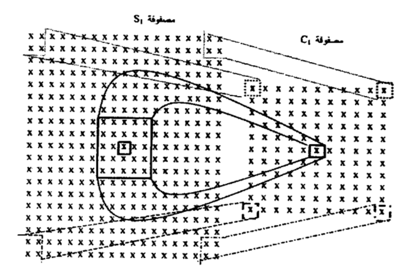
| | | | |
|---|---|---|---|
| $S_4^1, S_4^2$ | $\rightarrow C_4^1$ | $S_4^{10}$ | $\rightarrow C_4^6$ |
| $S_4^3, S_4^4$ | $\rightarrow C_4^2$ | $S_4^{11}, S_4^{12}$ | $\rightarrow C_4^7$ |
| $S_4^5$ | $\rightarrow C_4^3$ | $S_4^{13}$ | $\rightarrow C_4^8$ |
| $S_4^6, S_4^7$ | $\rightarrow C_4^4$ | $S_4^{14}$ | $\rightarrow C_4^9$ |
| $S_4^8, S_4^9$ | $\rightarrow C_4^5$ | $S_4^{15}, S_4^{16}$ | $\rightarrow C_4^{10}$ |

الآن سنعتبر حقل الاستقبال لوحدة في واحدة من المصفوفات (عند كل من المستويات المختلفة للشبكة). ترى وحدة واحدة في أي مصفوفة من مصفوفات الطبقة $S_1$ جزءاً حجمه 3×3 وحدة من نموذج الدخل؛ هذا يعني أن الوحدة $S_{1_{i,j}}^1$ تستقبل إشارات من وحدات الدخل $U_{i-1,j-1}$، $U_{i-1,j}$، $U_{i-1,j+1}$، $U_{i,j-1}$، $U_{i,j}$، $U_{i,j+1}$، $U_{i+1,j-1}$، $U_{i+1,j}$، $U_{i+1,j+1}$، و $S_{1_{i,j}}^2$ تستقبل إشارات من نفس وحدات الدخل التسعة.

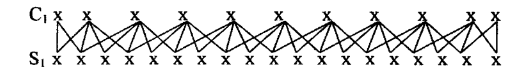
ترى وحدة في المصفوفة $C_1$جزءاً حجمه 5×5 وحدة من مصفوفة أو مصفوفتين $S_1$. وترى الوحدات في زاوية المصفوفة $C_1$ فقط جزءاً من المنطقة التي ستراها إذا توضعت في مركز المصفوفة، بسبب سقوط جزء من حقل رؤيتها خارج المصفوفة (المصفوفات) التي منها تستقبل إشاراتها، كما هو موضح في الشكل (5.11).

يحدث التنحيف أو الترقيق بسبب كون حجم كل مصفوفة $C_1$ أصغر من المصفوفة $S_1$. حقل رؤية وحدة $C_1$ موضح في الشكل (5.11)؛ يشير الرمز x إلى مكان توضع الوحدات. من المناسب مشاهدة مصفوفة $C_1$ كأنها متوضعة على قمة مصفوفة $S_1$ الموافقة.

تتوسع المصفوفة $C_1$ بعدئذ إلى ما وراء المصفوفة $S_1$، بحيث تستقبل وحدات زاوية المصفوفة $C_1$ إشارات من أربع وحدات فقط في مصفوفة $S_1$. من المناسب تلخيص المعلومات في الشكل (5.11) بالنظر إلى شريحة وحيدة البعد لنموذج ثنائي البعد كما هو موضح في الشكل (6.11).

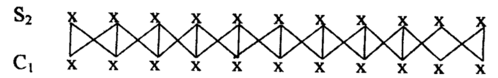



**الشكل 5.11**: الوصلات من مصفوفة واحدة S1 إلى وحدات من المصفوفة C1




**الشكل 6.11**: مقطع عرضي للوصلات من مصفوفة $C_1$ إلى مصفوفة $S_2$

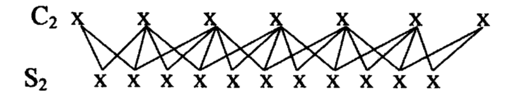
عند المستوى الثاني، كل وحدة $S_2$ ترى منطقة حجمها 3×3 وحدة من كل مصفوفة من مصفوفات$C_1$ الثماني. ولما كان حجم كل مصفوفة من مصفوفات $S_2$ هو 11×11 وحدة، وهو نفس الحجم كما في مصفوفات $C_1$، فلا يحدث ترقيق عند هذا المستوى. فقط وحدات $S_2$ التـــي لا تستقبل إشارات من وحدات $C_1$ التسع (في كل من مصفوفات $C_1$ الثماني) هي وحدات الزاوية في مصفوفات $S_2$؛ تستقبل وحدات الزاوية هذه إشارات من أربع وحدات $C_1$ فقط (في كل من مصفوفات $C_1$ الثماني). مخطط شريحة أحادية البعد موضح في الشكل (7.11).




**الشكل 7.11**: مقطع عرضي للوصلات من مصفوفة $C_1$ إلى مصفوفة $S_2$

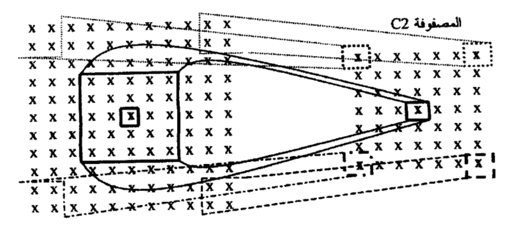
ترى وحدات $C_2$ منطقة حجمها 5×5 وحدة من مصفوفة $S_2$ (أو مصفوفات) التـــي تستقبل منها إشاراتها. حقل رؤية مصفوفة $C_2$ موضح في الشكل (9.11). سنرى عملية

الترقيق التـــي ستكون مشابهة لتلك التـــي حدثت في الطبقة الأولى. ثانية يلخص المخطط الأحادي البعد الموضح في الشكل (8.11) المعلومات.



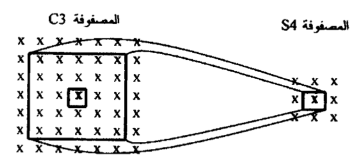

الشكل8.11: مقطع عرضي للوصلات من مصفوفة $S_2$ إلى مصفوفة $C_2$

كل وحدة $S_3$ ترى منطقة حجمها 3×3 وحدة من كل مصفوفة من 22 مصفوفة $C_2$؛ لا يحدث ترقيق عند هذا المستوى. كل مصفوفة $C_3$ ترى منطقة حجمها 3×3 وحدة من مصفوفة $S_3$ (أو المصفوفات) المتصلة معها، ومن ثم لا يحدث ترقيق في المستوى الثالث، لأن المصفوفات $C_3$ لها نفس حجم مصفوفات $S_3$ وهو 7×7 وحدة. ولما كانت وحدة $C_3$ بإحداثيات $i,j$ ترى منطقة من مصفوفة $S_3$ مركزة عند $i,j$، فإنه ليس ضرورياً رسم المخطط.



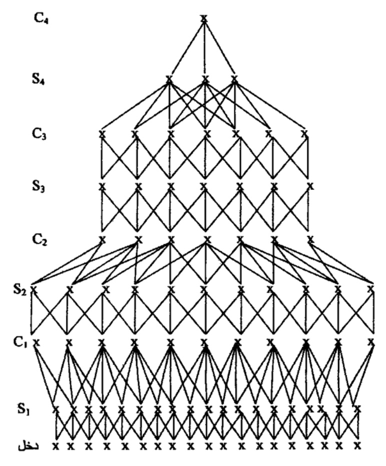

الشكل 9.11: الوصلات من مصفوفة $S_2$ إلى مصفوفة $C_2$

ترى كل مصفوفة من مصفوفات $S_4$ منطقة حجمها 5×5 وحدة من كل مصفوفة من مصفوفات $C_3$ الثلاثين. يوضح الشكل (10.11) حقل رؤية مصفوفات $S_4$. لاحظ أن تخفيض عدد الوحدات يحدث بين المستويين الثالث والرابع بدلاً من حدوثه ضمن المستوى كما في الحالة السابقة. أيضاً لاحظ أنه عوضاً عن تخطي الوحدات، فإن وحدات الزوايا الآن ستعالج بصعوبة (أهملها إذا أردت).




الشكل **10.11**: التوصلات من مصفوفة $C_3$ إلى مصفوفة $S_4$

كل مصفوفة $C_4$ هي فعلياً وحدة مفردة ترى مصفوفة (أو مصفوفات) $S_4$ بحجم 3×3 وحدة كاملة. يمكن تلخيص كل المعلومات فيما يخص نماذج الوصلات بين الوحدات في الطبقات المتنوعة بواسطة مخطط أحادي البعد، كما هو موضح في الشكل (11.11).

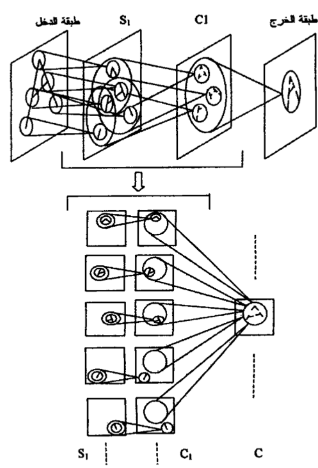



الشكل **11.11**: مقطع عرضي لنماذج توصيل شبكة النيوكونيترون [212]

الآن سنفترض أن الشبكة دربت على تعرف بعض الأشياء المعقدة كواحد من الأحرف الإنكليزية A , B, C. عندما يقدم أحد هذه الأحرف لطبقة دخل الشبكة وليكن الحرف A

ستصبح خلايا في عدة مصفوفات في $S_1$ فعالة. تعين خلايا الاستجابة بواسطة تركيب معالم المستوى المنخفض التـــي تركب الشيء والمحل النسبـــي الذي عنده يقدم الشيء إلى الدخل. تتفعل من خلايا هذه المصفوفات فقط التـــي تكون حساسة لمعالم المستوى المنخفض (المحتواة في الشيء). هذه المفاهيم موضحة في الشكل (12.11).

في الطبقة $S_1$ تكون معالم المستوى المنخفض مثل ∧ (رأس الحرف A) مكتشفة في واحدة من مصفوفات القمة كما هو موضح. تكتشف معالم أخرى في مصفوفات أخرى. عندما تضم هذه المعالم تنبه الخلايا في طبقات S الأعلى (التالية) حتـــى يكون في آخر الأمر الحرف بالكامل مميزاً في الطبقة $C_k$ الأخيرة.

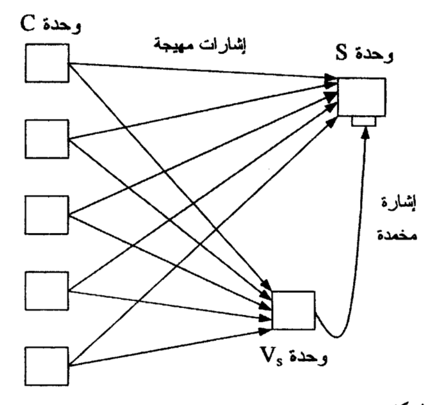



الشكل12.11: تعرُّف معالم المستوى المنخفض بواسطة عدة مصفوفات $S_1$

## 3.11 خوارزمية تدريب النيوكونيترون

إن إشارة خرج وحدة في مصفوفة نوع S (خلية في أي طبقة من طبقات S) هي تابع

لإشارات تهييج مستقبلة من وحدات في الطبقة السابقة ولإشارات تخميد مستقبلة من نفس تلك الوحدات. عموماً هناك نوعان من وحدات الطبقة S ونوعان من وحدات الطبقة C: مخمدة ومهيجة. لتوضيح ذلك سنرمز للوحدات المهيجة للطبقات S و C بـ $U_S$ و $U_C$ وللوحدات المخمدة بـ $V_S$ و $V_C$ على الترتيب. النظام التوصيلي للوحدات موضح في الشكل (13.11).




**الشكل 13.11**: وصلات من وحدات مصفوفة طبقة C إلى وحدات طبقة S

سنستخدم وحدة مساعدة V إشارة خرجها الذاهبة إلى الوحدة S متناسبة مع المعيار الإقليدي (المثقل) للإشارة المرسلة بواسطة وحدات الدخل.

سنعتمد الرموز التالية:

$c_i$ : الخرج من الوحدة C

$s_i$ : الخرج من الوحدة S

$v$ : الخرج من الوحدة V

$w_i$ : الوزن القابل للتعديل من الوحدة C إلى الوحدة S

$w_o$ : الوزن القابل للتعديل من الوحدة V إلى الوحدة S

$t_i$ : الوزن المجمد من الوحدة C إلى الوحدة V

$u_i$ : الوزن المجمد من الوحدة S إلى الوحدة C

تعطى الإشارة المرسلة بواسطة الوحدة المخمدة V بالعلاقة التالية:

$$v = \sqrt{\sum\sum t_i c_i^2} \qquad (1.11)$$

حيث أخذت المجاميع عبر كل الوحدات المتصلة مع V في أي مصفوفة وعبر كل المصفوفات.

عولجت طبقة الدخل كأنها المستوى C0. وهكذا، تكوِّن وحدة S نموذجية دخلها المعياري كما يلي:

$$x = \frac{1+e}{1+v\,w_o} - 1 \quad , \quad e = \sum_i c_i\, w_i \qquad (2.11)$$

حيث e دخل الشبكة المهيج من وحدات C، و $vw_o$ هو دخل الشبكة من الوحدة V.

وستكون إشارة الخرج:

$$S = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \qquad (3.11)$$

تعمل الإشارة المخمدة لمعايرة استجابة الوحدة S بأسلوب مشابه نوعاً ما للأسلوب المستخدم في خوارزمية ART2 (نظرية الطنين المتكيف، ستشرح في فصول لاحقة).

خرج وحدة الطبقة C هو تابع لدخل الشبكة المستقبل من كل الوحدات، في كل مصفوفات S، الذي يغذى إلى مدخل وحدة الطبقة C.

كما رأينا في وصف البنية، إن الدخل هو نموذجياً من 9 أو 25 وحدة في كل مصفوفة S أو مصفوفتين S أو ثلاث مصفوفات S.

دخل الشبكة هو:

$$c(net) = \sum_i s_i\, u_i \qquad (4.11)$$

وسيكون الخرج:

$$c=\begin{cases}\dfrac{c(net)}{a+c(net)} & c(net)>0\\ 0 & c(net)\le 0\end{cases} \tag{5.11}$$

يعتمد الوسيط a على المستوى ويكون 0.25 للمستويات 1 و2 و3 ويكون 1 للمستوى الرابع.

تجري عملية تدريب النيوكونيترون طبقة تلو الأخرى. تكون الأوزان من وحدات C إلى وحدة S معدلة وملائمة، وكذلك تكون الأوزان من الوحدة V إلى الوحدة S، أما الأوزان من الوحدات C إلى الوحدة V فتكون مثبتة. تثبت أيضاً الأوزان من مصفوفة طبقة S إلى مصفوفة طبقة C الموافقة. وتكون هذه الأوزان أقوى للوحدات التـــي هي أقرب ولكن ليس هناك مقياس متري خاص معين.

كمثال لنوع نموذج الوزن الذي يمكن أن يستعمل، اعتبر:تابع مسافة تكون فيه المسافة من وحدة الطبقة S، $S_{i-k,j-h}$، إلى وحدة الطبقة C، $C_{i,j}$، تساوي $|h|+|k|$. تعطى مصفوفة الأوزان الممكنة للوزن من $S_{i-k,j-h}$ إلى $C_{i,j}$ كما يلي:

$$u(S_{i-k,j-h},C_{i,j})=\frac{1}{1+|k|+|h|} \tag{6.11}$$

لمنطقة وصل بحجم 5×5 وحدة، وهي الأوزان الواصلة من الطبقة $S_2$ إلى الطبقة $C_2$ ستكون بالقيم التالية:

| | | | | |
|---|---|---|---|---|
| 1/5 | 1/4 | 1/3 | 1/4 | 1/5 |
| 1/4 | 1/3 | 1/2 | 1/3 | 1/4 |
| 1/3 | 1/2 | 1 | 1/2 | 1/3 |
| 1/4 | 1/3 | 1/2 | 1/3 | 1/4 |
| 1/5 | 1/4 | 1/3 | 1/4 | 1/5 |

ونموذج الأوزان سيكون نفسه لكل وحدة $C_2$.

الأوزان المثبتة من وحدات C إلى وحدات V المخمدة تكون أيضاً متناقصة انسياباً كتابع للمسافة. والأوزان إلى وحدات الطبقة S (من وحدات الدخل أو من وحدات الطبقة C في الطبقة السابقة) دربت على التتالي. الأوزان من وحدات الدخل إلى الوحدات $S_1$ دربت

وجمدت.

تستمر العملية مستوى تلو الآخر حتـــى يتم الوصول إلى طبقة الخرج.وسنصف الآن عملية التدريب بالتفصيل.

**تدريب الأوزان من وحدات الدخل إلى وحدات الطبقة $S_1$:**

دربت كل مصفوفة من 12 مصفوفة في الطبقة $S_1$ لتستجيب لنموذج دخل 3×3 مختلف. نماذج التدريب لكل المصفوفات في الطبقة $S_1$ موضحة في الشكل (4.11). تستجيب كل وحدة في المصفوفة $S_1^1$ لنموذج (قطعة مستقيمة أفقية) عندما يظهر في جزء مصفوفة الدخل التـــي تستقبل الوحدة الخاصة إشاراتها منها. إن نموذج الأوزان إلى كل الوحدات في $S_1^1$ هو نفسه.

لتدريب كل الوحدات في المصفوفة $S_1^1$، لدينا فقط وحدة واحدة للتدريب (المسماة، وحدة المركز)؛ فنموذج التدريب للمصفوفة $S_1^1$ يمثل بواسطة مركز مصفوفة الدخل (والإشارة المنشودة ترسل إلى مصفوفات $S_1$ لتعيّن أن وحدة مركز المصفوفة $S_1^1$ هي الوحدة المدربة). الأوزان من وحدة الدخل $U_{i+k,j+h}$ إلى وحدة المصفوفة $S_1^1$، $S_{1i,j}^1$، ستكون معدلة كما يلي:

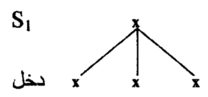
$$\Delta w\left(U_{i+k,j+h}, S_{1i,j}^1\right) = \alpha t\left(U_{i+k,j+h}, S_{1i,j}^1\right) c_{i+k,j+h}$$

لأول طبقة S، الإشارة $c_{i+k,j+h}$ هي ببساطة إشارة دخل. الوزن $t\left(U_{i+k,j+h}, S_{1i,j}^1\right)$ هو وزن مجمد للوحدة المخمدة. وهكذا، يكون الوزن المعدل متناسباً مع الإشارة المستقبلة بواسطة الوحدة المخمدة. يعدل الوزن من الوحدة المخمدة إلى الوحدة S بالكمية:

$$\Delta w_o = \alpha c_{i,j}$$

حيث القيم الأولية للأوزان المعدلة تساوي الصفر، وأعطي معدل التعليم $\alpha$ قيمة كبيرة نسبياً بحيث تتعلم الوحدة S المدربة استجابتها المنشودة بعد بضع تمثيلات للنموذج فقط. عندما تصبح أوزان وحدة المركز معينة، تعطى كل الوحدات الأخرى في المصفوفة $S_1^1$ تماماً نفس قيم الأوزان. بهذه الطريقة، تُدرَّب الوحدة المركزية لتكون فعالة عندما يقدم نموذج دخل في مركز حقل الدخل، ولكن الوحدات الأخرى في المصفوفة $S_1^1$ تستجيب لنفس نموذج الدخل (في هذه الحالة، قطعة مستقيمة أفقية) عندما يظهر في أجزاء أخرى من حقل

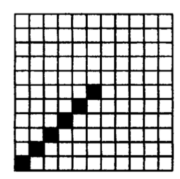
الدخل. بأسلوب مشابه، تُدرَّب الوحدة المركزية للمصفوفة $S_1^2$ لتستجيب لنموذج دخل معطى كما هو موضح في الشكل (4.11). بعد تعيين قيم الأوزان، تعطى كل الوحدات الأخرى في هذه المصفوفة نفس القيم. يستمر التدريب بنفس الطريقة لـــ 12 مصفوفة في الطبقة $S_1$. يظهر مخطط المقطع العرضي للحقل المستقبل للوحدة $S_1$، الموضح في الشكل (14.11) سبب كون نماذج التدريب لهذا المستوى هي 3×3 فقط؛ هذا يعنـــي أن كل الوحدة يمكن أن ترى.




**الشكل 14.11:** مقطع عرضي للحقل المستقبل للوحدة S1.

### تدريب الأوزان من وحدات $C_1$ إلى وحدات $S_2$:

تستقبل وحدة المركز في كل مصفوفة من الطبقة $S_2$ إشارات من تسع وحدات في كل واحدة من مصفوفات $C_1$. تُدرَّب كل مصفوفة $S_2$ لتستجيب لعدد صغير من النماذج. مثلاً، يمكن أن تكون نماذج التدريب للمصفوفة $S_2^4$ نماذج عديدة عن النموذج البسيط المبين في الشكل (15.11).

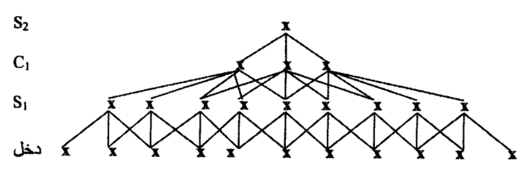


**الشكل 15.11:** نموذج تدريب بسيط لمصفوفة $S_2^4$.

كما ذكرنا سابقاً، يقدم نموذج التدريب إلى مركز حقل الدخل، والوحدة المركزية في المصفوفة $S_2^4$ تخصص لتتعلم النموذج. تعدل أوزان الوصلات التسعة من كل مصفوفة من مصفوفات $C_1$ باستعمال نفس معدل التعليم كما هو الحال في الطبقة الأولى.

لاحظ بوجه عام، أن عدداً قليلاً جداً من مصفوفات $C_1$ سيستجيب لإشارة الدخل، بحيث لن يكون نموذج الوصل الفعلي(بدون أوزان صفرية) من مستوى $C_1$ إلى مستوى $S_2$ موسعاً كما أشير في الوصف العام. ومع أن نماذج التدريب هذه هي قطع مستقيمة قطرية، فإن نماذج تدريب مصفوفات $S_2$ أخرى تحقق تراكيب من نماذج التدريب البسيطة التـــي دربت عليها من قبل $S_1$ و$C_1$ لتعطي استجاباتها. كما هو الحال في الطبقة الأولى، بعد تدريب الوحدة المركزية على نماذج تدريبها (نموذجياً، بتغيرات أربعة على نموذج التدريب الأساسي)، ستثبت وحدات أخرى في تلك المصفوفة أوزانها عند نفس القيم لأوزان الوحدة المركزية.

يجري تدريب كل مصفوفة في الطبقة $S_2$ بنفس الأسلوب. عندما تكون كل المصفوفات مدربة، تثبت الأوزان ونشرع بتعديل أوزان المستوى التالي. إن مخطط المقطع العرضي للحقول المستقبلة، الموضح في الشكل (16.11) يبين العلة في كون نماذج التدريب لهذا المستوى هي 11×11. إذا اقتفينا عكسياً أثر الوصلات من الوحدة المركزية عند المستوى $S_2$ إلى مستوى الدخل، سنرى أن منطقة 11×11 في مستوى الدخل تؤثر في مصفوفة المستوى $S_2$.




**الشكل 16.11:** حقل الاستقبال للوحدة المركزية في مصفوفة المستوى $S_2$

### تدريب الأوزان من وحدات $C_2$ إلى وحدات $S_3$:

يجري تدريب مصفوفات المستوى $S_3$ باتباع نفس الإجراءات للمستويات الأدنـــى (التـــي أوزانها الآن مثبتة). حقل الاستقبال للوحدة المركزية هو الآن مصفوفة الدخل كاملة،

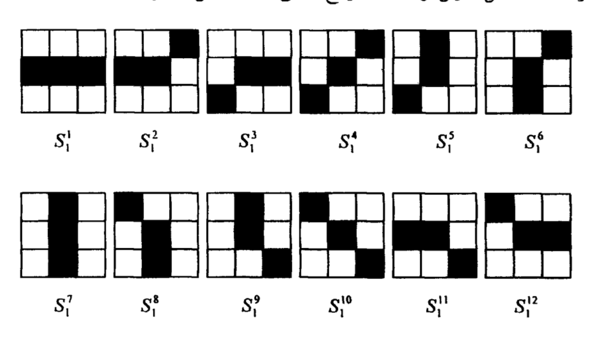
بحيث تكون نماذج التدريب 19×19.

**تدريب الأوزان من وحدات $C_3$ إلى وحدات $S_4$:**

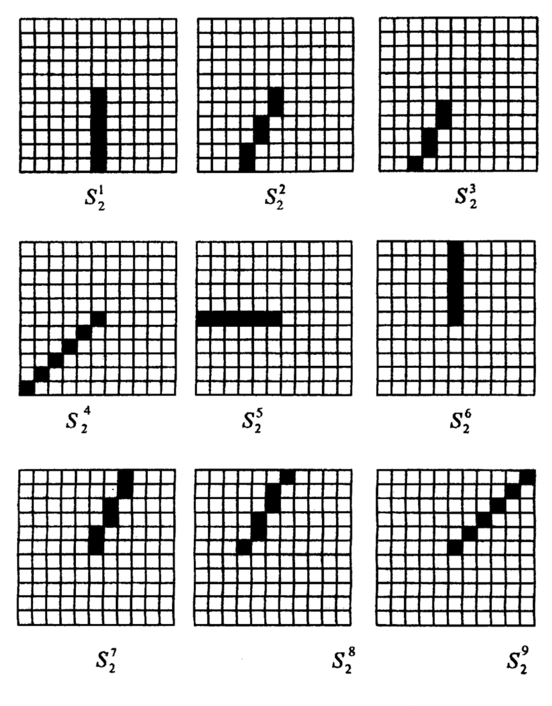
التدريب النهائي للأوزان، لوحدات $S_4$ الست عشرة مبنـــي على نماذج بسيطة متنوعة معطاة فيما يلي.

**عينة من نماذج تدريب الطبقة S [212]:**

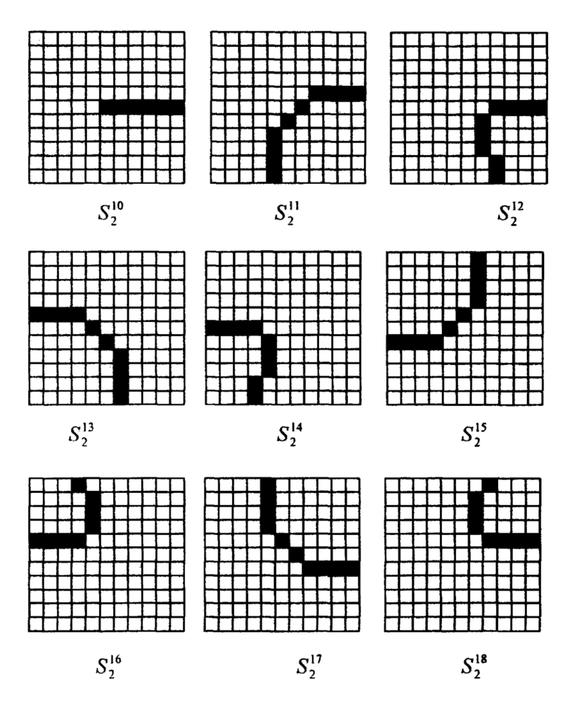
عينة مـــن نماذج التدريب لمصفوفات الطبقات Sموضحة في الأشكال (17.11) حتـــى (20.11). ونماذج تدريب مصفوفات $S_2$ مبينة في الشكل (18.11). يبين الشكل (19.11) والشكل (20.11) عينة نماذج تدريب كلٍّ من المصفوفات عند المستوى $S_3$ والمستوى $S_4$؛ نموذجياً، استعمل تغييرين أو ثلاثة للنموذج المعطى لتدريب كل مصفوفة.



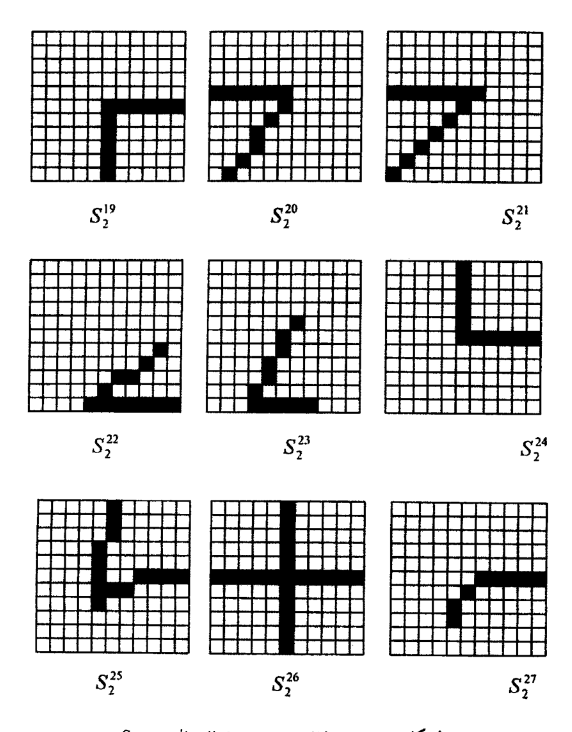

**الشكل 17.11**: نماذج تدريب الطبقة $S_1$ لشبكة نيوكونيترون



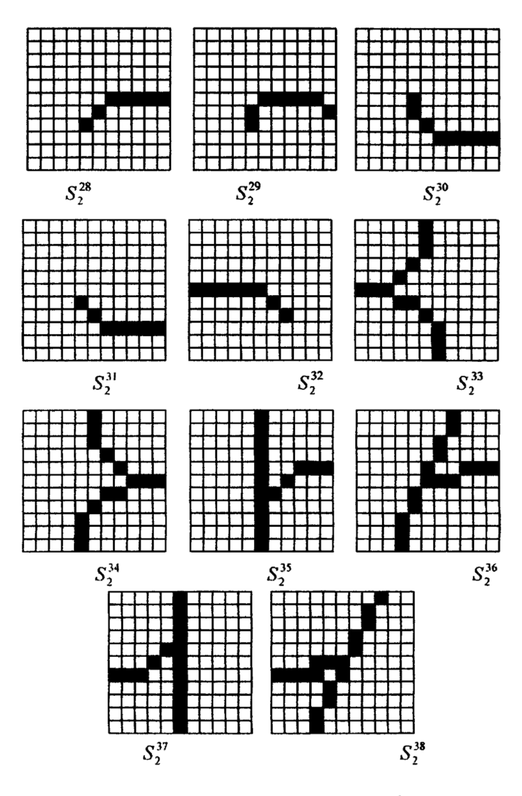

**الشكل 18.11** (أ) نماذج تدريب مصفوفات المستوى S2



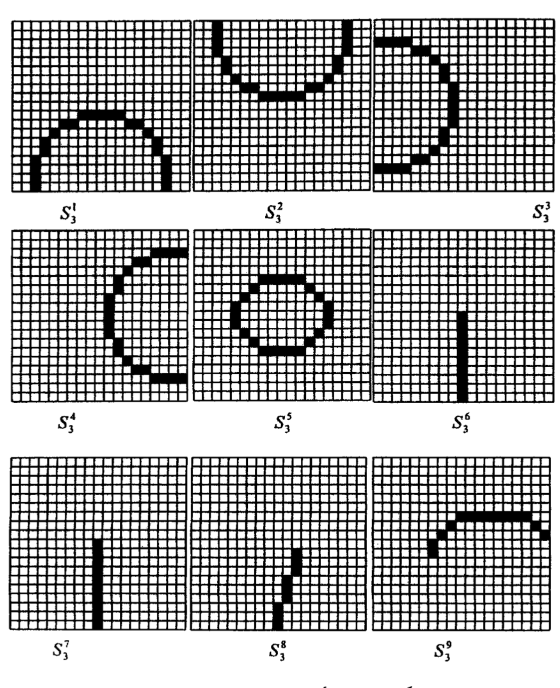

**الشكل 18.11** (ب) نماذج تدريب مصفوفات المستوى $S_2$



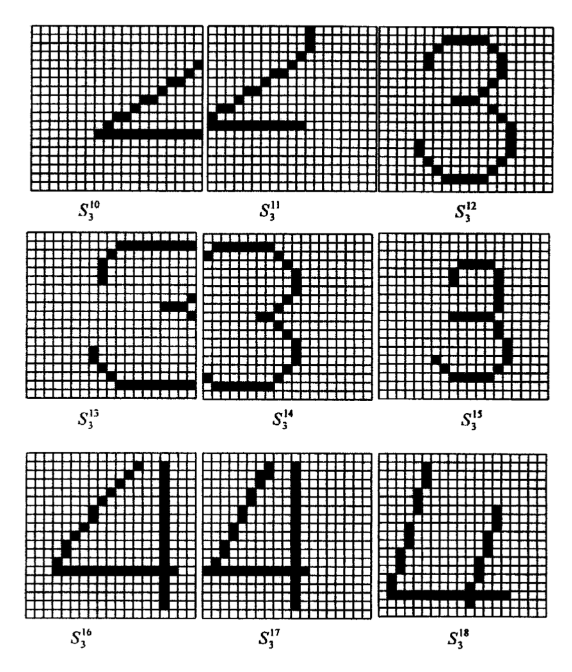

**الشكل 18.11** (ج) نماذج تدريب مصفوفات المستوى $S_2$



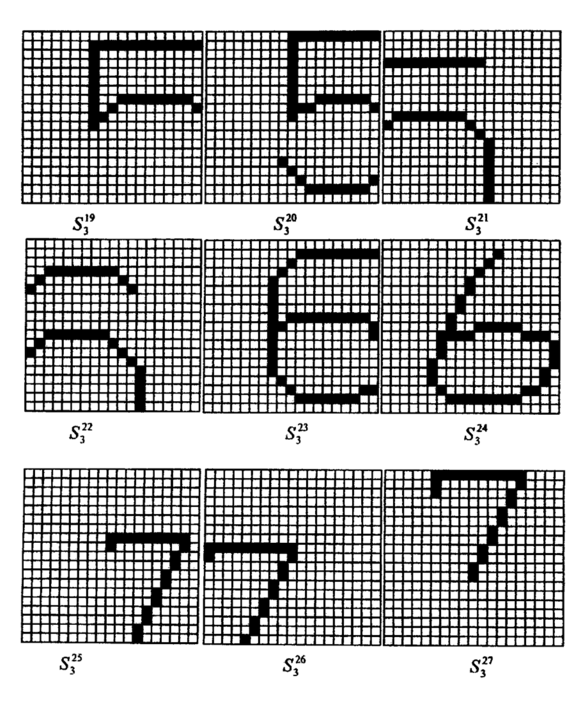

**الشكل 18.11** (د) نماذج تدريب مصفوفات المستوى S2



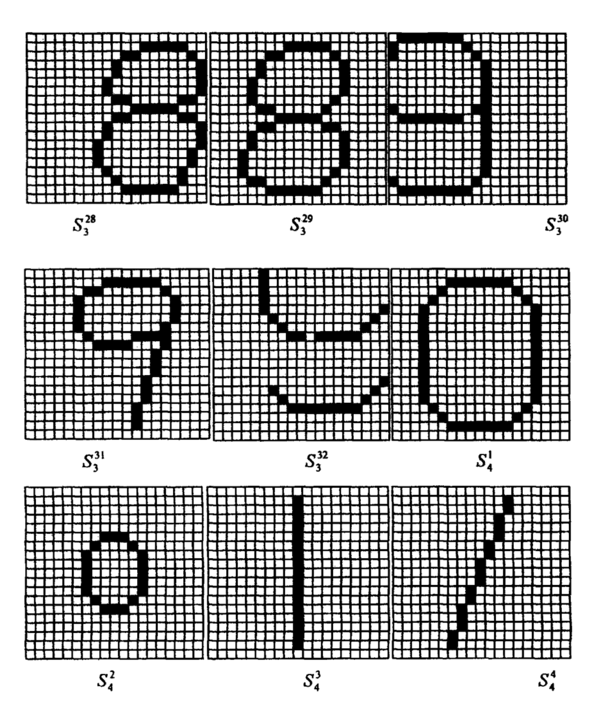

**الشكل (19.11)** (أ) نماذج تدريب مصفوفات المستوى $S_3$



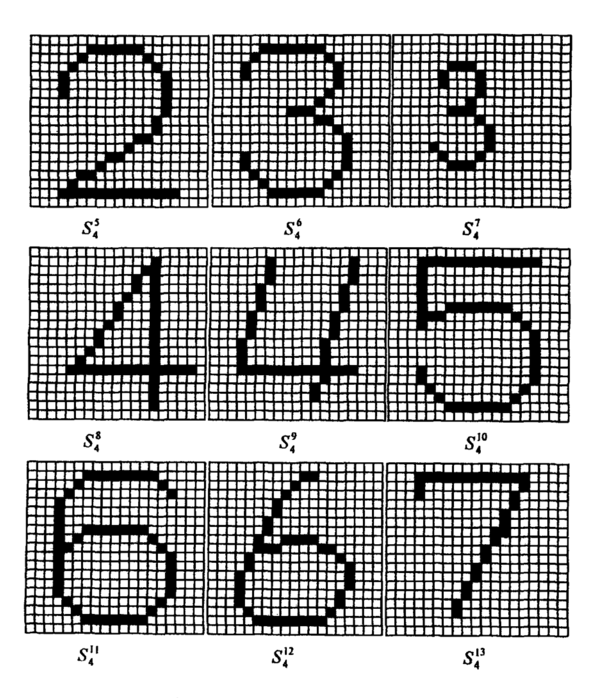

**الشكل (19.11)** (ب) نماذج تدريب مصفوفات المستوى $S_3$



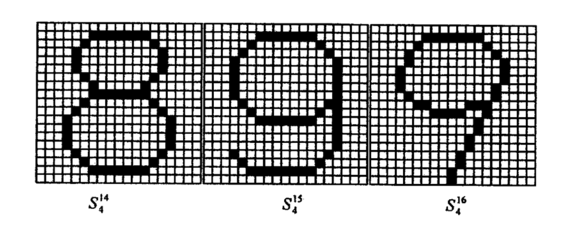

**الشكل (19.11)** (ج) نماذج تدريب مصفوفات المستوى $S_3$




**الشكل (20.1)** (أ) نماذج تدريب مصفوفة المستوى $S_4$، وتتمة الشكل (19.11) (ج) نماذج تدريب مصفوفة المستوى $S_3$




**الشكل (20.1)** (ب) نماذج تدريب مصفوفات المستوى S4



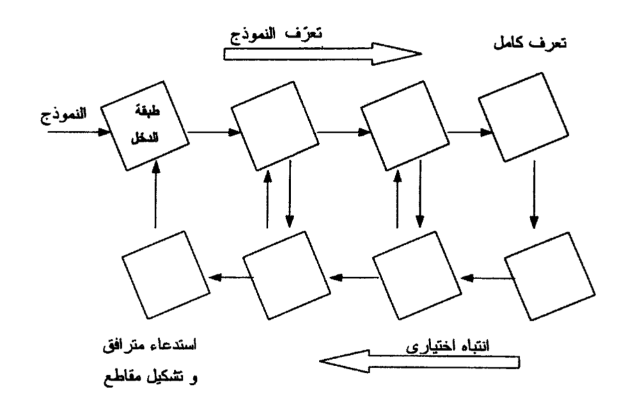

**الشكل (20.1)** (ج): نماذج تدريب مصفوفات المستوى S4

## 4.11 شبكات النيوكونتيرون المعززة

دمجت في النسخ المعززة لشبكة النيوكونتيرون ممرات تغذية عكسية من طبقة الخرج إلى الطبقات السابقة لتسهيل استدعاء الترافق الذاتـــي والتعرف المتتابع على أشياء عديدة في نفس الصورة.

تنفذ إشارات التغذية العكسية نوعاً من عملية تركيز الانتباه، حيث يمكن أن يشكل من الشيء المتعلم لوحده مقاطع ويعرف، ومن ثم لا يستجيب النظام فيما بعد لهذا الشيء. بعدئذ يزاح الانتباه عن الشيء الأول وتستجيب الشبكة للشيء الثانـــي، وهكذا حتـــى يتم تعرف كل شيء ضمن الصورة على التتالي. وكذلك ننجز هذا النوع من تركيز الانتباه الاختياري عندما نراقب أشياء متعددة في حقل رؤيتنا.

سنركز الآن على شيء واحد في الصورة وسنصرف الانتباه عن الأشياء الأخرى. وبأسلوب مشابه لشبكة النيوكونتيرون الأساسية، تنفذ ممرات التغذية الأمامية في الشبكة المعززة وظيفة تعرُّف الشكل أو النموذج، أما ممرات التغذية العكسية فإنها تنفذ الوظائف الإضافية مثل الانتباه الاختياري، والاستدعاء المترافق، وتشكيل مقاطع الشيء. يوضح الشكل (21.11) جميع ممرات الإشارات في شبكة النيوكونتيرون المعززة.

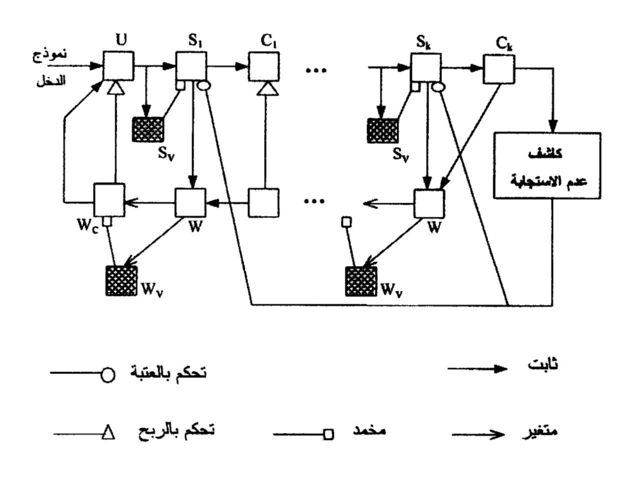



**الشكل21.11:** ممرات تدفق الإشارة في شبكة نيوكونتيرون معززة

يوضح الشكل (22.11) جزءاً من ممر وحيد للإشارة بين الخلايا، من خلايا المصفوفات الأولية والنهائية، حيث وضحت ممرات التغذية الأمامية والعكسية على هذا الشكل.

سنرمز لخلايا التمييز في الطبقة C النهائية بـ $C_k$، أما الخلايا المشار إليها بـ $W_s$، $W_c$، $W_v$ فهي خلايا ممرات التغذية العكسية التي تؤدي دوراً مشابهاً للخلايا S وC وV في الممرات الأمامية على الترتيب. إن للخلايا W أوزاناً ثابتة ومتغيرة على وصلات دخلها، وتقارن إشارات الخرج المخمدة والمهيجة مع خلايا S وC وV الموافقة في الممرات الأمامية بأدوار متبادلة للخلايا $W_c$ و$W_s$.

بعد تعرف الشيء، يعاد خرج طبقة التمييز، واحداً من خلايا $C_k$، عكسياً إلى الطبقات الأدنى من خلال ممرات التغذية العكسية مرحلة بعد مرحلة حتى الوصول إلى طبقة الاستدعاء. تسلك إشارات التغذية العكسية نفس الطريق العكسي كما في حالة إشارات التغذية الأمامية. يمكن تحقيق هذا بواسطة استجابة الخلايا في ممرات التغذية الأمامية التي تعمل عندئذ كبوابة لإشارات التغذية العكسية.




**الشكل22.11**: الوصلات بين الخلايا في ممرات الإشارة الأمامية والعكسية

في هذه الطريقة، تتدفق إشارات التغذية العكسية مستهدية طريقها بواسطة إشارات التغذية الأمامية بحيث تصل إلى نفس محل البداية كإشارة التغذية الأمامية. ولما كانت إشارات التغذية العكسية تنتشر للخلف بواسطة إشارة الخلية C المتنشطة في طبقة الخرج، فإن مركبات الإشارة الموافقة للنموذج المميز فقط تصل إلى طبقة الاستدعاء. ومن ثم تعزز إشارات طبقة الاستدعاء عملية تعرف الشكل، وهذا ما يساعد على استدعاء ترافق ذاتـــي وعملية تشكيل مقاطع الشيء.

إن أوزان ممرات التغذية العكسية مشابهة لأوزان ممرات التغذية الأمامية، والنوعان يكونان مقويين معاً خلال عملية التدريب، (فعلياً، الخلايا $W_S$ لها أوزان ثابتة والخلايا $W_C$ لها أوزان متغيرة لتسهيل عملية فتح وإغلاق الإشارة).

يراقب كاشف عدم الاستجابة في الشكل (22.11) خرج خلايا التمييز. عندما تُكتشَف حالة عدم الاستجابة، ترسل هذه الوحدة إشارات للخلايا S الكاشفة للملمح في كل المراحل

لتخفيض قيمة عتبتها وجعلها أكثر حساسية للملامح في طبقة الدخل. وهذا يساعد على تفعيل، خلية معرفية واحدة على الأقل في الخرج.

عندما يقدم نموذجان أو أكثر إلى طبقة دخل الشبكة في آن واحد، يمكن أن تصبح خليتان في الخرج أو أكثر مبدئياً فعالتين في البداية. الكل ماعدا واحداً سيوقف الاستجابة حالاً، تبعاً لوصلات التخميد الجانبية المتنافسة بين كل خلايا الطبقة S في ممرات التغذية الأمامية. عندما توقف كل الخلايا استجابتها ماعدا خلية واحدة، تصل مركبات الإشارة العكسية الموافقة للخلية المتنشطة فقط طبقة الاستدعاء. هذه الإشارات تؤدي إلى عملية تشكيل مقاطع نموذج الإشارة حتـــى ولو كانت نسخة مشوهة عن النموذج الأصلي للتدريب. أيضاً تزيد إشارات التغذية العكسية ربح إشارات الممر الأمامي لتقوية الانتباه الاختياري المركز على نموذج واحد، ويؤدي التفسير اللحظي لإشارة التغذية العكسية إلى قطع الانتباه الاختياري لنموذج آخر. هذا تبعاً لضياع ربح الإشارة في خلايا الطبقة C التـــي تتطلب تسهيل الدعم من خلايا $W_C$ في الممر العكسي. وهكذا تصبح مجموعة أخرى من خلايا الدخل وخلايا الطبقة C فعالة ويجري تمييز نموذج آخر في الصورة وتشكيل مقاطعه. نتيجة تكرار هذه العملية، يمرر الانتباه أو يمنع لكل من النماذج المعلمة في الصورة، حتـــى عندما تتشابك بعضها مع بعض نوعاً ما.

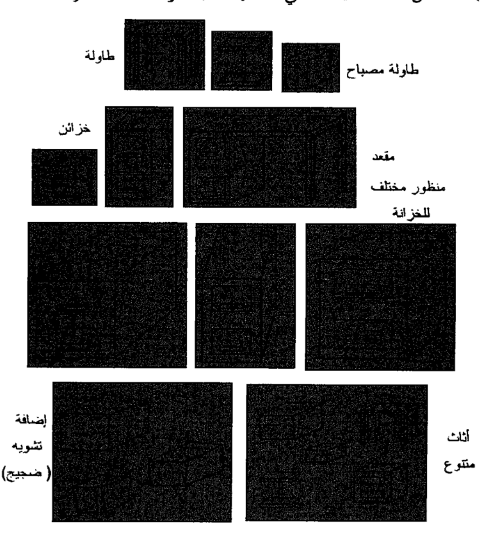
## 5.11 تطبيقات شبكة النيوكونتيرون

ارتبطت معظم التطبيقات المنشورة لشبكة النيوكونتيرون بالبصريات أو تعرف الأشكال بما في ذلك تمييز أحرف الكتابة اليدوية وأنواع متنوعة من صور الأشياء. سنناقش هنا تطبيقان لهذه الشبكات.

### 1.5.11 تعرف أنواع زوايا وصل الأشياء

طور Lee وPatterson عام 1991 [86] شبكة نيوكونتيرون هجينة قادرة على تمييز نماذج شكلت سلكياً لأنواع مختلفة من أثاث مكتب، حيث ركبت هذه الأشياء بواسطة خطوط

مستقيمة ووصلات، مثل خزانة الملفات، والطاولات، ومقعد المصباح، وهكذا. يوضح الشكل (23.11) أمثلة عن هذه الأشياء التـــي تعلمتها الشبكة واستطاعت تمييزها.

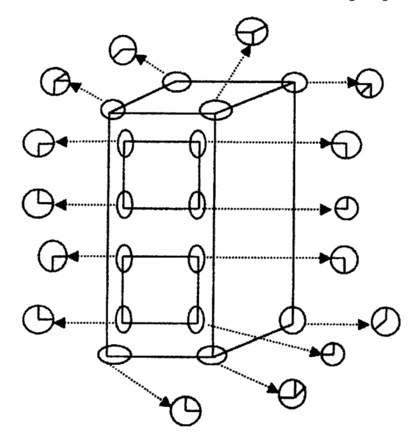



**الشكل23.11**: أمثلة عن أثاث مكتب تعرَّفتها الشبكة

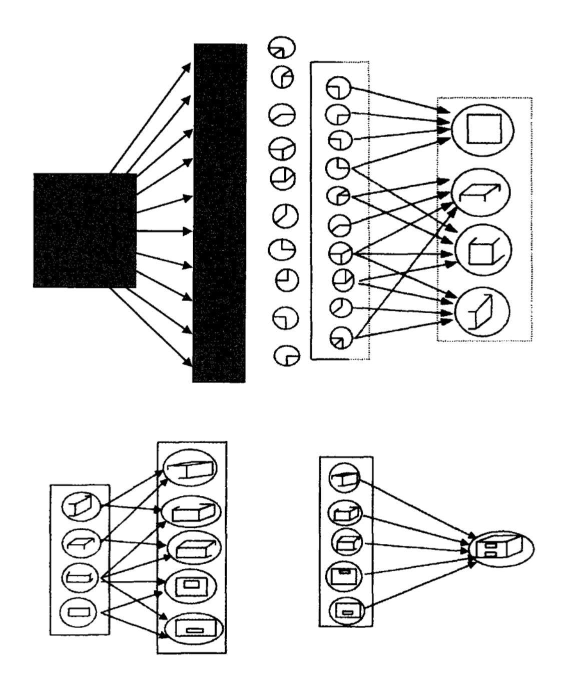
إن ملامح المستوى المنخفض التـــي ميزت من قبل الشبكة هي الوصلات وتُعدّ الأوليات البنيوية للأشياء، وقد استعمل حوالي 28 وصلة لتركيب ملامح المستوى المنخفض. علمت الوصلات واستجيب لها بواسطة أول طبقتين، ثم رُكّب بعضها مع بعض لتشكيل المناطق في الزوج الثانـــي من الطبقات. ركبت المناطق في الزوج الثالث من الطبقات لتشكيل مستويات فراغية مغلقة للأشياء الجزئية، ويستجيب الزوج الأخير من الطبقات للأشياء

كاملة. هذه العمليات التركيبية موضحة في الشكل (24.11) حيث جرى تمييز خزانة الملفات تمييزاً كاملاً.

تألفت طبقة الدخل من 128×128 عنصر صورة، كما أن للطبقات المتتالية وصلات أقل مع وجود طبقة نهائية أنشئت خلال طور التدريب عندما يجري تعلم شيء جديد. بسبب مقدرة النيوكونتيرون على تمييز الإزاحة والتشويه للأشياء غير التغيرية، فإن الشبكة تستطيع تمييز نفس الشيء عند اتجاهات مختلفة عن الاتجاه الأول الذي دربت عليه، كما أنها تتعامل مع مستوى عالٍ من ضجيج الخلفيات. وقد تحقق بعض النجاح أيضاً في التجارب المنفذة على الأشياء الظاهرة جزئياً.



**الشكل (24.11)** (أ): تعرف ملمح إلى شيء بواسطة الطبقات المتتالية



الشكل (24.11)(ب): تعرف ملمح إلى شيء بواسطة الطبقات المتتابعة

## 2.5.11 تعرف أحرف الكتابة اليدوية

لقد استعمل باحثون كثر شبكات النيوكونتيرون لتعرف الأحرف اللاتينية والأرقام العربية المكتوبة يدوياً أو المطبوعة بالآلة الكاتبة. كانت توصيلات الشبكة على النحو التالي:

| الطبقة | عدد المصفوفات في الطبقة | حجم كل مصفوفة من الوحدات |
|---|---|---|
| دخل (U) | 1 | 19 ×19 |
| $S_1$ | 12 | 19 × 19 |
| $C_1$ | 8 | 21 × 21 |
| $S_2$ | 80 | 21 × 21 |
| $C_2$ | 33 | 13 × 13 |
| $S_3$ | 97 | 13 × 13 |
| $C_3$ | 64 | 7 × 7 |
| $S_4$ | 47 | 3 × 3 |
| $C_4$ | 35 | 1 × 1 |

طور هذه الشبكة Fukushima وWake عام 1991 [185]، حيث كانت قادرة على تمييز 35 حرفاً مكتوباً يدوياً، وهي الأرقام العربية 0,1,2,...,8,9 و25 حرفاً من اللغة الإنكليزية مع استبعاد الحرف O. كانت المجموعة المكانية الصغيرة للخلايا من طبقة الدخل U إلى الطبقة S1 مؤلفة من مصفوفة مربعة بحجم 3×3 وحدة، وكذلك كانت المجموعات من الطبقة C إلى الطبقة S مصفوفة مربعة بحجم 7×7 و5×5 و3×3 وحدة على الترتيب. كان العدد الكلي للخلايا 70045 خلية(خلايا S وC وV والخلايا المستقبلة)، استعملت لتعرُّف 36 حرفاً أبجدياً في هذا التطبيق.

دربت هذه الشبكة بواسطة التعليم بمعلم. حيث يجري التعليم من طبقات المستوى الأدنـــى إلى طبقات المستوى الأعلى، وكما في حالة التعليم بدون معلم يجري تدريب الطبقات العليا فقط بعد أن تكون الطبقات السفلى قد دربت تدريباً كاملاً.

وضعت كل الأوزان المعدلة بقيم بدائية صفرية. يختار المعلم بعدئذ مصفوفة الخلية S لتكون مدربة. يقدم نموذج التدريب إلى طبقة الدخل U ويختار المعلم خلية ضمن المصفوفة المختارة لتعمل كخلية أصل (أو بذرة). يشار إلى الخلية الأصل بواسطة مكان مركز حقلها المستقبل. تعزز الأوزان المتغيرة للخلية البذرة لتصبح مستجيبة للوصلات المناسبة المعينة بواسطة الملامح التـــي يجب أن تعلم. تتناسب كمية التعزيز لكل وصلة للخلية البذرة مع شدة الاستجابة للخلية التـــي تقاد منها الوصلة المناسبة.

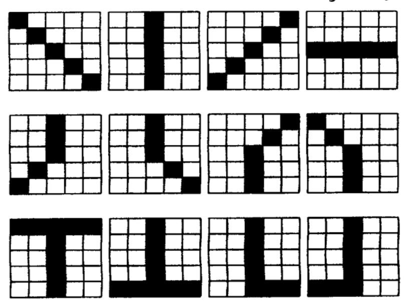
كنتيجة لمبدأ التعليم هذا، توضع الأوزان لتنمو بحيث تشكل طبقة معايرة تلائم تماماً

التوزع المكانـــي لاستجابات الخلايا في الطبقة السابقة. في هذه الطريقة، تكتسب الخلية الأصل إمكانية استنباط ملمح إشارات الدخل خلال التدريب. تتبع جميع الخلايا الأخرى في المصفوفة الخلية الأصل كالنمو الكريستالي، أي تقوى أوزانها لتتبع الخلية البذرة، ومن ثَم، سيكون لها نفس التوزع المكانـــي مثل خلية الأصل. وهكذا تتعلم كل الخلايا في المصفوفة استخراج نفس الملمح لكن عند أمكنة مختلفة.

دربت الطبقة $S_1$ على استخراج مركبات خطية بسيطة باتجاهات مختلفة مشابهة للنماذج الموضحة في الشكل (25.11) استعمل 12 نموذجاً فقط لتدريب 12 مصفوفة $S_1$، وقدمت كل نماذج التدريب للشبكة مرة واحدة فقط.

دربت الخلية عند مركز مصفوفة الخلية لتنتخب دائماً كخلية أصل.

باعتبار أن الحقل المستقبل لكل خلية في هذه المصفوفة هو 3×3 وحدة، فإن مساحة 3×3 فقط مركزية لكل نموذج تدريب ستكون فعالة خلال التدريب. سيكون لخلايا الطبقة $S_2$ حقولاً مستقبلة حجمها 9×9 وحدة.



**الشكل25.11**: ملامح مستوى منخفض نموذجية مكتشفة في مصفوفات الطبقة $S_1$

هناك 80 مصفوفة في هذه الطبقة توافق تراكيب مختلفة للملامح الأولية البسيطة المستعملة لتدريب الطبقة $S_1$.

دربت الطبقة $S_3$ لتشكل استجابة لملامح كلية أكثر (ملامح معقدة أكثر). تضم هذه

الطبقة الملامح من الطبقة $S_2$ باستعمال 97 مصفوفة خلية.

دربت الطبقة النهائية $S_4$ لتمييز الأحرف تمييزاً كاملاً. لهذه الطبقة 47 مصفوفة خلية، واستعملت نسخ مختلفة لمعظم الأحرف لتدريب الطبقة على تعلم لا تغيري بتشويه. أحد المحاسن الرئيسة للتعليم بمعلم المستعمل في هذه الشبكة هو زمن التدريب الذي كان قصيراً جداً. بعد تحديد مجموعة التدريب، لزم فقط 13 دقيقة لتدريب الشبكة على محطة SUN PARC، بالمقارنة مع الشبكات المتعددة الطبقات الأمامية التغذية التـــي دربت لتفعل نفس المهمة بالانتشار الخلفي والتـــي تطلبت ثلاثة أيام زمن تدريب (le Cun عام 1990[84]).

# الشبكات المبنية احتمالياً
# Stochastic-based Networks

ينفّذ هذا النوع من الشبكات العصبونية الذي سندرسه في هذا الفصل نمذجةً احتمالية، لكنها لا تعمل بالأسلوب الاحتمالي كما هو الحال في شبكات آلة بولتزمان التـــي درسناها في الفصل التاسع.

لقد رأينا أمثلة عديدة عن الشبكات التـــي تبنـــي نماذج احتمالية لوسطها المحيط. مثلاً، تبنـــي الشبكات المتعددة الطبقات الأمامية التغذية بخوارزمية تدريب الانتشار الخلفي نماذج التراجع بدون معاملات غير خطية. سنتناول في هذا الفصل الاستعمال المباشر لطرائق التقدير بدون معاملات (nonparametric estimation methods) في بناء الشبكة.

ستكون عملية التدريب مباشرة، وتبنـــي الشبكة جيداً، وستكون سريعة جداً أيضاً، بالإضافة إلى ذلك لن نواجه هنا مشكلة الأصغر المحلي مطلقاً.

يُنظر إلى هذا النوع من الشبكات على أنه استمرار لنوع الشبكات الموصوفة في الفصل العاشر، باعتبار أن أصناف الشبكات المدروسة هنا هي أيضاً ذاتية النمو، حيث تنمو بتوليد طبقة (تطبيق) داخلية خلال عملية التعليم بمعلم، وذلك بإضافة عقدة لكل نموذج تدريب أو لكل تجمع من النماذج.

## 1.12 تمهيد

لسنوات عديدة خلت، كانت التقنيات الإحصائية المبنية على نظرية الاحتمال الكلاسيكية هي التقريب المقبول لمهام التصنيف وتعرف الأشكال، حيث كانت هذه الطرائق المبنية على التراجع الخطي (الانحدار الخطي)، والتجمعات، والارتباط، ومصنفات بايز (Bayes) قياسات مقبولة، عموماً، وشائعة في مسائل التصنيف والتمييز وتعرف الأشكال.

لقد رأينا من قبل حالات عديدة تنفّذ فيها الشبكات العصبونية الصنعية أنواعاً معينة من مسائل تعرف الأشكال من خلال نوع من التحليل الإحصائي. في هذا الفصل سنوسع هذه الطرائق لتشمل بنيتين للشبكات العصبونية تنفذان نمذجة احتمالية لنموذج الوسط المحيط، لكن بطريقة مباشرة. في كلا الشبكتين، استعملت عينات نماذج التدريب لإيجاد نماذج تقدير بدون معاملات لتوزيعات العينات.

ركبت الشبكات من خلال إضافة عقد داخلية تمثل الخواص الإحصائية لنماذج التدريب، ثم ركب خرج هذه العقد ليعطي تطبيقاً منشوداً للدخل إلى الخرج.

تعمل كلتا الشبكتين بالأسلوب التعييني وليس الاحتمالي كما في حالة آلة بولتزمان التي درسناها من قبل، وتشتركان بنفس النموذج التوصيلي والبنية. وهاتان الشبكتان اقترحتهما الباحث Donald Specht وهو أحد طلاب Widrow بين أعوام 1988[214] وحتى 1990 [145].

## 2.12 الشبكة العصبونية الاحتمالية

### Probabilistic Neural Network(PNN)

اقترح Specht الشبكة العصبونية الاحتمالية بين عامي 1988[214] و1990[145]. وقد بنيت هذه الشبكة على المفاهيم الكلاسيكية المستعملة في مسائل تعرف الأشكال وتمييزها، وخاصة، نماذج الشبكة العصبونية الصنعية لمصنف بايز (Bayes) الشائع (Mood وGraybill عام 1962[215])، تلك التقنية التي تجعل المخاطر المتوقعة للتصنيف الخاطئ للنماذج أصغرية. يمكن أن توصف عملية مصنفات بايز كما يلي: ليكن x شعاع دخل ذي بعد n يميز الأشياء التي تنتمي لواحد من الصفوف الممكنة K، ولتكن $f_1(\mathbf{x}), f_2(\mathbf{x}),...,f_k(\mathbf{x})$ هي توابع كثافة الاحتمال لمجتمعات الصفوف K على الترتيب. ولتكن $p_1, p_2, \cdots, p_k$ الاحتمالات الاستنتاجية حيث ينتمي شعاع الملمح **x** إلى الصف الموافق. وما نريده الآن هو تابع القرار (الفصل) $\mathrm{d}(\mathbf{x}) = C_i$، $i = 1,2,\cdots,K$، بحيث يصنف **x** على أنه منتمٍ إلى الصنف $C_i$ انتماء أفضل وفقاً لحاسة ما.

مثلاً، يمكن أن نعرف هذا الشكل الأفضل للتصنيف كتابع يصنف نماذج الدخل بمخاطرة

أصغرية للتصنيف غير الصحيح. ولتكن $L_1, L_2, \cdots, L_k$ توابع الضياع المرافقة للقرار الخاطئ بحيث يحدث الضياع عندما:

$$x \in C_j \text{ و } d(\mathbf{x}) = C_i \text{ ،} i \neq j \qquad (1.12)$$

ويساوي الضياع الصفر في حالة القرار الصحيح. قاعدة قرار بايز لهذا النوع من مسائل التصنيف تقارن بين الجداءات الناتجة:

$$p_1 L_1 f_1(\mathbf{x}), p_2 L_2 f_2(\mathbf{x}), \ldots, p_k L_k f_k(\mathbf{x}) \qquad (2.12)$$

ومن ثم تختار الصف الموافق لقيمة الجداء الأكبر. وهكذا، إذا كان:

$$p_i L_i f_i(\mathbf{x}) > p_j L_j f_j(\mathbf{x}) \qquad (3.12)$$

في حالة $j(i \neq j) = 1, 2, \cdots, K$

فإن قاعدة القرار تخصص $\mathbf{x}$ للصف $C_i$ وليس $C_j$.

يمكن أن تستعمل أيضاً تغيرات معيار الاختيار، بما في ذلك استعمال تابع الكلفة أو تابع آخر لمعاقبة اختيار التصنيف غير الصحيح.

أحد الانتقادات الرئيسية لتقنيات تصنيف بايز هو فقدان المعلومات حول توزيعات احتمالات الصفوف. هناك عادةً مجاهيل يجب أن تقدر بطريقة ما؛ فالاحتمالات الاستنتاجية $p_i$ يمكن أن تعرف وتقدر بسهولة مباشرة من عينة أشعة النماذج، لكن توابع الاحتمال $f_i(\mathbf{x})$ تكون صعبة التقدير.

بالطبع يمكن افتراض شكل توزيع ما (مثلاً توزيع طبيعي) وبعدئذ تقدير الوسطاء غير المعروفة باستعمال تقنيات إحصائية قياسية.

يُعدّ استعمال طرائق التقدير بدون معاملات أكثر مناسبة في حالة عدم وجود أي معرفة حقيقية. إحدى التقنيات بدون معاملات القوية هي تلك المبنية على استعمال نوافذ بارزن (Parzen) (Parzen عام 1962[216]). يعطى الشكل العام للمقدِّر بالمعادلة التالية:

$$f_n(x) = \frac{1}{n\lambda} \sum_{i=1}^{n} \varphi\left(\frac{x - x_i}{\sigma}\right) \qquad (4.12)$$

حيث $x_i$ متحولات عشوائية موزعة توزيعاً منتظماً ومستقلة بتابع توزيع مستمر استقلالاً مطلقاً. تابع التثقيل (Weighting) $\varphi$ يجب أن يكون محدداً ومحققاً للشروط التالية:

$$\int_{-\infty}^{\infty}|\varphi(y)|\,dy < \infty$$
$$\lim_{y\to\infty}|\varphi(y)| = 0 \qquad (5.12)$$
$$\int_{-\infty}^{\infty}\varphi(y)\,dy = 1$$

والتابع $\sigma = \sigma(n)$ يجب أن يختار بحيث:

$$\lim_{n\to\infty}\sigma(n) = 0 \quad , \quad \lim_{n\to\infty} n\sigma(n) = \infty \qquad (6.12)$$

أثبت بارزن أن هذه المقدرات (estimators) منسجمة (متماسكة)، وتتقارب تدريجياً إلى توزيع أساسي عند نقاط العينات، وذلك عندما تكون ناعمة ومستمرة.

وُسِّعت نتائج بارزن أيضاً إلى توزيع متعدد المتحولات من قبل **Cacoullons** عام **1966** .[217]

إحدى الأشكال المفيدة للتابع المثقل $\varphi$ هو تابع غوص (Gauss)الأسي المتعدد المتحولات. في هذه الحالة، تأخذ المعادلة (4.12) الشكل التالي:

$$f_i(\mathbf{x}) = \frac{1}{2\pi^{n/2}}\frac{1}{k_i}\sum_{j=1}^{k_i}\exp\left[-\frac{(\mathbf{x}-\mathbf{x}_{ij})^T(\mathbf{x}-\mathbf{x}_{ij})}{2\sigma^2}\right] \qquad (7.12)$$

حيث n بعد المتحول العشوائي، و $k_i$ عدد نقاط عينات النموذج في الصف $C_i$، و $\mathbf{x}_{ij}$ هي العينة رقم j المنتمية للصف $i$، والعامل هو وسيط النعومة الذي يجب أن يحدد تجريبياً.

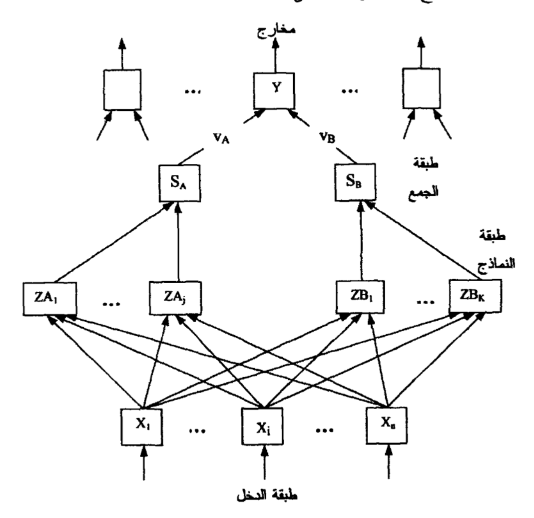
هذا المُقدِّر الخاص هو مجموع أو متوسط $k_i$ حداً أسيّاً (توابع غوص)، حيث خصص حد واحد لكل عينة في الصف المعطى. يحدد شكل الحدود الأسية المجموعة بواسطة ثابت التنعيم $\sigma$.

تعطي قيمةٌ كبيرة لـــ $\sigma$ شكلاً لمقدِّر بارزن هو منحنٍ بهضاب مسطحة ذات انحدارات ناعمة وغير حادة، على حين تعطي قيمة صغيرة لـــ $\sigma$ شكلاً لمقدر بارزن هو منحنيات شوكية ضيقة شديدة الانحدار. إن اختيار $\sigma$ سيؤثر في خطأ التقدير، لكن بوجه عام، لا يتأثر الإجراء كثيراً نتيجة التغيرات في قيمة $\sigma$ (كما ذكر **Specht** عام **1988**[214]).

اقترح **Specht** بنية الشبكة العصبونية بين عامي **1988** و**1990** لتأدية عمل مصنف بايز باستعمال مقدِّر نافذة بارزن (المعادلة 7.12)) وذلك لتقدير توزيعات الاحتمال لعينات الصف.

يسمح استعمال التابع المثقل $\varphi$ من نوع بارزن (أو غيره) الأسي كتوابع تفعيل للشبكة الاحتمالية أن تتعلم بناء حدود الفصل (القرار) غير الخطية التـــي تعتبر تقريباً لأسطح قرار مصنف بايز الأمثلي.

بنية الشبكة الاحتمالية المبسطة موضحة في الشكل (1.12)، حيث تتألف من أربع طبقات: طبقة الدخل وطبقة النماذج وطبقة الجمع وطبقة الخرج. هذه البنية المبسطة المعطاة في الشكل (1.12) تضم فقط وحدتين في طبقة الجمع ووحدة في طبقة الخرج، أي إن الشبكة الموضحة في هذا الشكل تُعدّ مصنفاً ثنائياً (الصف A أو الصف B). في حالة مصنفات متعددة الصفوف (ليس اثنان فقط)، سيكون للشبكة أكثر من وحدة خرج واحدة، وسيكون لها أيضاً وحدات جمع إضافية واحدة لكل صف.




**الشكل 1.12:** بنية الشبكة العصبونية الاحتمالية

ستجعل كل نماذج الدخل بطولٍ واحديٍّ قبل المعالجة. استعملت طبقة الدخل لتوزيع نماذج الدخل على وحدات طبقة النماذج التالية لها، وستكون وحدات طبقة النماذج متصلة اتصالاً كاملاً مع طبقة الدخل من خلال أوزان قابلة للتعديل.

تضم طبقة النماذج K وحدة، حيث خُصِّصت وحدة واحدة لكل نموذج تدريب، ووضعت قيم أوزان الوحدات في هذه الطبقة مساوية لنماذج التدريب المختلفة، ثم جُعلتْ أشعة الوزن معيارية بطول يساوي الواحد. تنفِّذ وحدات طبقة النماذج جداءً نقطياً على شعاع نموذج الدخل وشعاع وزن الوحدة. ولما كان هذان الشعاعان معياريين بطول يساوي الواحد، فإن هذا الجداء يكافئ عملية الجداء التالية:

$$(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)$$

هذا التعبير له نفس شكل الأس في المعادلة (7.12).

إن توابع التفعيل لوحدات طبقة النماذج هي من النوع الأسي المشابه للتوابع المستعملة في المعادلة (7.12). يصبح الجداء السابق الذي يقيس المسافة بين نموذج الدخل $\mathbf{x}$ وشعاع وزن الوحدة $\mathbf{w}_i$ أسياً بواسطة وحدات النماذج قبل تمرير خرجها إلى طبقة الجمع. وهكذا فإن خرج الوحدة رقم j في طبقة النموذج يعطى بما يلي:

$$f(\mathbf{x},\mathbf{w}_j)=\exp\left[\sum_{i=1}^{k_i}\frac{(\mathbf{x}_i-\mathbf{w}_{ij})}{2\sigma^2}\right] \qquad (8.12)$$

تتصل مخارج طبقة النماذج اختيارياً مع وحدات طبقة الخرج بالاعتماد على صف النماذج التـــي تمثلها.

تجمع وحدات الجمع مخارج طبقة النماذج لإتمام حساب المعادلة (7.12). ولما كانت كل وحدة جمع تجمع المداخل من وحدات النماذج لنفس الصف فقط، فإن مخارجها ستكون تقديراً لتابع كثافة احتمال الصف المعطى بواسطة المعادلة (7.12)، حيث يكون التابع $\varphi$ مركزاً عند عينات التدريب ($\mathbf{w}_i$).

تعطي وحدات الخرج إشارة خرج ثنائية، حيث يكون لكل منها وصْلتا دخل. تقوم هذه الوحدات بحساب جداء خرج وحدات الجمع وعامل الوزن $C_k\,(k=1,2)$ :

$$C_k = -\frac{p_2 L_2}{p_1 L_1}\frac{k_1}{k_2} \qquad (9.12)$$

حيث $k_1$ و $k_2$ هما أرقام النماذج من الصفين 1 و2 علـى الترتيب، (مقارنة مـع الشكل (1.12) فإن $v_A=1$ و $v_B=C_k$).

من المحاسن الرئيسية للشبكات العصبونية الاحتمالية السرعة التـي يمكن أن تدرب بها. حيث لا تلزم إجراءات متكررة أو ممرات تغذية عكسية في عملية التدريب.

أثبتت دراسة مقارنة للشبكة الاحتمالية مع الشبكة المتعددة الطبقات الأمامية التغذية باستعمال خوارزمية الانتشار الخلفي (MLFF with BP)، أن الشبكة الاحتمالية تدربت أسرع بـ 200000 مرة من شبكة الانتشار الخلفي (Specht عام 1990[145]).

التطبيق الذي استخدم في هذه الدراسة المقارنة كان مسألة تصنيف جسم سفينة، مع ملاحظة أن إنجاز كلتا الشبكتين كان متكافئاً.

يلزم في تدريب الشبكة الاحتمالية تكرار واحد فقط لمجموعة التدريب. وهكذا، يمكن أن تعدل حدود القرار (الفصل) في الزمن الحقيقي باستعمال معطيات جديدة عندما تصبح متوفرة. أيضاً، تعتبر الشبكات الاحتمالية متسامحة مع العينات الضجيجية، وقد تعمل مع عينات مبعثرة أيضاً، ويمكن أن تعمل بإحصائيات زمن متغير، لأن النماذج القديمة يمكن أن تكون مكتوبة بنماذج جديدة.

أخيراً، يمكن تصميم بنية الشبكة عتادياً (Hardware) باستعمال عصبونات صنعية تعمل على التوازي.

لا تستعمل الشبكات الاحتمالية لمهام التصنيف فقط، بل يمكن أن تستعمل لتقدير الاحتمالات الاستنتاجية؛ أي احتمالات أن ينتمي نموذج الدخل **x** إلى الفئة C، $P[C|x]$، ويمكن أن تُستعمل كذواكر مترافقة، وفي مسائل تطبيق عامة أخرى.

مثلاً، عندما تُستعمل كذاكرة مترافقة، تكون فئة النموذج الناقص (الكامل جزئياً) هي الدخل وتُغيَّر المداخل غير المعروفة لتحديد أو تعريف القيم التـي تجعل خرج الشبكة أعظمياً.

من المساوئ الرئيسية للشبكات الاحتمالية أن حجم طبقة النماذج فيها يمكن أن ينمو نمواً

كبيراً جداً عندما تكون مجموعة التدريب المستعملة ضخمة (تذكّر أن كل نموذج تدريب يؤدي إلى إضافة عقدة نموذج في طبقة النماذج بواسطة خوارزمية التدريب).

للحد من هذه المشكلة وتهذيبها، يمكن أن تحلّ نماذج الصفوف الأولية محل المجموعات الضخمة من النماذج الفردية، حيث توفر هذه النماذج الأولية مقدرات تمثيلية لاحتمالات المجموعات.

في مجموعة التجارب التـــي نفذها Burrascano عام 1991[187]، استعمل تعليم التكميم الشعاعي LVQ (Learning Vector Quantization)، (سيشرح بالتفصيل في الفصل التالي)، لإنقاص عدد عقد طبقة النماذج عشر مرات، أو مئة مرة عند استعمال أشعة رموز تعليم التكميم الشعاعي لتقريب توزيعات احتمالات المجموعات. ولكن نتائج إنجاز التصنيف كانت أقل دقة بمقدار 2 إلى 3% فقط في حالات ذات عقد أقل. لذا، فإن هذا البديل أو الخيار سيستعمل فقط عندما يتوفر لدينا مجموعة معطيات ضخمة جداً.

## 3.12 تعليم الشبكة الاحتمالية Training Probabilistic Network

تبنـــى الشبكة العصبونية الاحتمالية خلال تنفيذ خوارزمية التدريب. تمثل كل مجموعة من الوحدات في طبقة النماذج واحداً من الصفين اللذين تنتمي إليهما نماذج التدريب، وتوافق كل وحدة نموذج واحدة (ضمن مجموعة وحدات كل صف) نموذج تدريب واحد.

أولاً تجعل نماذج التدريب معيارية بطول يساوي الواحد ويوضع شعاع الوزن لوحدة النموذج $ZA_j$ (الشكل(1.12))بقيمة شعاع التدريب رقم j الذي ينتمي إلى الصف A.

عندما يقدم نموذج تدريب، تضاف وحدة نموذج جديدة موافقة للصف الصحيح إلى الشبكة، وتوضع قيمة أوزانها، وتوصل هذه الوحدة الجديدة إلى وحدة الجمع الصحيحة الموافقة.

ستكون خطوات الخوارزمية كما يلي:

1. في كل نموذج دخل تدريب **x(p)**، حيث p =1, 2, .., P كرر الخطوات 2 و3
2. إحداث (إضافة) وحدة نموذج $Z_p$: شعاع الوزن لهذه الوحدة **w(p) = x(p)** (الوحدة $Z_p$ ستكون وحدة ZA أو وحدة ZB).

3. توصيل وحدة النموذج إلى وحدة الجمع:

**إذا كان** $\mathbf{x}(p)$ ينتمي إلى الصف A عندئذ توصل وحدة النموذج $Z_p$ (وحدة ZA) بوحدة جمع $S_A$.

**وإلا** توصل وحدة النموذج $Z_p$ (وحدة ZB) بوحدة الجمع $S_B$.

لتطبيق هذه الخوارزمية أولاً سنقوم بجعل نماذج الدخل معيارية بطول يساوي الواحد. في حالة أشعة واحدية معيارية، يمكن كتابة حد المجموع في العلاقة الأسية (8.12) كما يلي:

$$\exp\left[\frac{z(net_j)-1}{\sigma^2}\right]$$

وستكون خوارزمية تصنيف النماذج (المعيارية بطول يساوي الواحد) كما يلي:

1. إعطاء الأوزان قيماً أولية

2. في كل نموذج دخل مطلوب تصنيفه، كرر الخطوات من 3 حتـــى 5

3. في وحدات النماذج:

يحسب دخل وحدة النموذج المضافة:

$$z(net_j) = \mathbf{x}.\mathbf{w}_j = \mathbf{x}^T.\mathbf{w}_j$$

ويحسب خرجها كذلك:

$$z = \exp\left[\frac{z(net_j)-1}{\sigma^2}\right]$$

4. في وحدات الجمع:

اجمع المداخل من وحدات النموذج المتصلة بالوحدة، وحدة الصف B فقط سيضرب دخلها بـــ:

$$C_k = v_B = -\frac{p_2 L_2}{p_1 L_1}\frac{k_1}{k_2}$$

5. وحدة (القرار/الفصل) الخرج:

تجمع وحدة الخرج الإشارات من $S_A$ و$S_B$. سيصنف شعاع الدخل في الصف A إذا كان الدخل الكلي لوحدة الخرج موجباً.

الاحتمالات الاستنتاجية للصف A وللصف B ستكون نموذجياً نسبة عدد نماذج التدريب في الصف A إلى عدد نماذج التدريب في الصف B. في هذه الحالة فإن:

$$\frac{p_2}{p_1}\frac{k_1}{k_2}=1$$

ويصبح تعبير $v_B$ المبسط هو:

$$v_B=-\frac{L_2}{L_1}$$

تعتمد هذه النسبة على أهمية القرار (الفصل) وليس على إحصائيات الوضع، وإذا لم يكن هناك سبب لانحياز القرار(الفصل)، نأخذ $v_B = 1$.

## 4.12 الشبكة العصبونية التراجعية المعممة

## Generalized Regression Neural Network (GRNN)

استُعملت طرائق التنبؤ الإحصائي على نطاق واسع لحل مسائل عديدة منها التنبؤ بطقس الغد القريب، ومبيعات منتج جديد سيطرح في الأسواق، ومؤشرات مالية مختلفة مثل مؤشر تجارة الأسهم، واستهلاك الطاقة الكهربائية المتنامي مع مرور الوقت، وكذلك مسائل أخرى للتنبؤ في حقول متنوعة.

يشبه تحليل التراجع (هناك من يسمي التراجع بالانحدار) الإحصائي طرق الارتباط، حيث يُستعمل، عموماً، أداة للتنبؤ.

استعمل التحليل التراجعي لإلباس (fitting) منحنٍ ناعم بعدد من نقاط عينة المعطيات التـــي تمثل ظاهرة متغيرة باستمرار. إن تقنية الإلباس بمنحنٍ يمكن أن تستعمل للتنبؤ بقيم متحول أو متحولات على أساس المعلومات المتوفرة من قياسات على متحولات (إيضاحية) مستقلة أخرى.

لإنجاز التحليل التراجعي، يجب أن يختار نموذج تابعي يمكن أن يمثل العلاقة بين المتحولات المستقلة وغير المستقلة.

يأخذ النموذج الخطي العام (التراجع الخطي وهناك من يسميه الانحدار الخطي) الشكل التالي:

$$z_k = X_k \beta + \varepsilon_k, \ k = 1,2, ...,P \qquad (10.12)$$

حيث هناك P مراقبة على المتحول غير المستقل $Z_k$ والمتحولات المستقلة $x_k=(x_{k1},x_{k2},...,x_{km})$، و$\beta$ شعاع غير معروف بُعده $m\times 1$ يمثل الوسيط الذي يجب أن يقدر، والمتحول العشوائي $\varepsilon_k$ هو توزيع غير مراقب.

النموذج الخطي المعطى في العلاقة (10.12) هو نموذج احتمالي من المفترض الاحتفاظ به لاستخدامه في مسائل تنبؤ عديدة.

في إنجاز التحليل التراجعي، يجب أن تكون الوسطاء التي تعرف العلاقة التابعية مقدرة باستعمال معيار إحصائي ما. مثلاً، لإنجاز تراجع خطي (انحدار خطي) بسيط بين متحولين $x$ و $y$، يفترض شكل تابعي هو خط مستقيم، ينبغي بعد ذلك تقدير وسطاء الميل وتقاطع المستقيم (مع المحاور).

إذا كان المتحولان معروفين أو متوقعين لهذه العلاقة الخطية، يمكن عندها تمثيل هذه العلاقة بواسطة مجموعة معادلات خط مستقيم، $y = a x + b$، حيث b هي نقطة التقاطع مع المحور y وa هو ميل المستقيم.

يمكن أن نحصل على تقديرات الوسطاء a وb من مقادير أزواج عينات مقاسة $(x_i, y_i), i, j = 1,2,..,n$ للمتحولين $x$ و$y$.

إن معيار التقدير الشائع هو المربعات الصغرى، أي إيجاد تلك القيم لوسطاء المستقيم الذي يجعل مجموع الفروق المربعة بين نقاط معطيات المراقبة والخط التراجعي بقيمة صغرى، كما هو موضح في الشكل (2.12). تعطى الوسطاء b وa بطريقة المربعات الصغرى بالعلاقات التالية:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n\left(\sum_{i=1}^{n} x_i^2\right) - \left(\sum_{i=1}^{n} x_i\right)^2} \qquad (11.12)$$

$$a = \frac{1}{n}\left[\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i\right] \qquad (12.12)$$

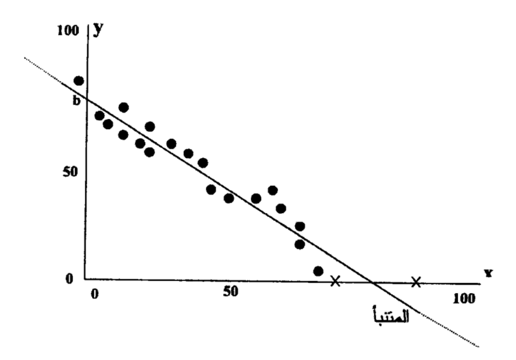
بعد أن يحصل على تقدير الوسطاء، يمكن استعمال الخط الملبس (المتنبأ به) للتنبؤ بقيم

المتحولات غير المستقلة لأية قيم جديدة للمتحولات المستقلة.

ينجز هذا من خلال الاستيفاء الداخلي (التوليد) (Interpolation) بين نقاط العينات أو من خلال الاستيفاء الخارجي (Extrapolation) للمستقيم إلى ما وراء نقاط معطيات العينة الأصلية. مثلاً، التنبؤ بواسطة الاستيفاء الخارجي موضح في الشكل (2.12) بالخط المتقطع الممتد مع الخط الملبس.

يشمل تحليل التراجع، عموماً، استعمال متحولات مستقلة وغير مستقلة متعددة وطرق تسوية العلاقات غير الخطية فيما بين التغيرات.

أحد عيوب طريقة التراجع هو ضرورة افتراض شكل تابعي بين المتحولات المستقلة وغير المستقلة، وهذا التابع غير معروف في كثير من المسائل. مع ملاحظة أن الافتراض غير الصحيح لهذا الشكل التابعي يمكن أن يؤدي إلى تنبؤات غير محققة للأغراض المرجوة.




**الشكل 2.12**: تقدير وسطاء الميل والتقاطع لمستقيم باستعمال معيار المربعات الصغرى

إن افـتـراض علاقة خطية بسيطة كتلك المعطاة في المعادلة (10.12) سيؤدي إلى وصول غير مضمون إلى النتائج المنشودة. لذا سنهتم فيما يلي، بوجه عام، بتنفيذ تحليل تراجع غير خطي.

ليكن $f(\mathbf{x},z)$ تابع كثافة الاحتمال المشترك لمتحول عشوائي شعاعي $\mathbf{x}$ ومتحول عشوائي

سلّمي z. يعرف تراجع z على x كمتوسط شرطي لـــ z مع x معطى، أي $E[z|\mathbf{x}]$، حيث

$$E[z|\mathbf{x}] = \frac{\int_{-\infty}^{\infty} z\, f(\mathbf{x}, z) dz}{\int_{-\infty}^{\infty} f(\mathbf{x}, z) dz} \qquad (13.12)$$

تكون الكثافة المشتركة f(x,z) عادة غير معروفة، لذا يجب أن تقدر من نقاط معطيات العينة لـــ z و x.

هناك الكثير من التقنيات المتوفرة لتقدير f(x,z) باستعمال قياسات على z وx، وقد أسست تقنية بدون معاملات شائعة على طريقة نافذة بارزن عام 1962 المشار إليها سابقاً. يعطى النوع الأول لمقدِّر نافذة بارزن في تابع الكثافة هذا بواسطة:

$$f_p(\mathbf{x}, z) = \frac{1}{2\pi^{(n+1)/2}\, \sigma^{n+1}} \frac{1}{P} \sum_{i=1}^{P} \left[ -\frac{(\mathbf{x}-\mathbf{x}_i)^T (\mathbf{x}-\mathbf{x}_i)}{2\sigma^2} \right] \exp\left[ -\frac{(\mathbf{z}-\mathbf{z}_i)^2}{2\sigma^2} \right] \qquad (14.12)$$

حيث P عدد نقاط العينة، وn بعد شعاع نقاط العينة $\mathbf{x}_i$، وx وسيط، و $\sigma$ ثابت النعومة.

من المعروف أن هذا المقدر منسجم، لذا فهو يميل لأن يكون متقارباً لتوزيع أساسي تدريجياً كلما ازداد عدد نقاط العينة P تحت شروط غير محصورة مرتبطة باستمرارية التوزيع الأساسي و$\sigma$.

يمكـــن تفسيـــر المقدر بأنه متوسط احتمالات P عينة، كل منها بعرض $\sigma$ لنقاط العينة $\mathbf{x}_i$ و$z_i$. إذا عوض المقدر $f_p(\mathbf{x}, z)$ في المعادلة (11.12)، فيمكن أن نحصل على شكل أقرب لتقدير التراجع المشار إليه بـــ $\hat{z}$. وقد أعطى النتيجة Specht عام 1991[218] بالعلاقة التالية:

$$\hat{z} = \frac{\sum_{i=1}^{P} z_i \exp\left( -\frac{D_i^2}{2\sigma^2} \right)}{\sum_{i=1}^{P} \exp\left( -\frac{D_i^2}{2\sigma^2} \right)} \qquad (15.12)$$

حيث يعرف العامل $D_i^2$ كما يلي:

$$D_i^2 = (\mathbf{x} - \mathbf{x}_i)^T (\mathbf{x} - \mathbf{x}_i)$$

يمكن أن تكون المتحولات z و $\hat{z}$ بقيم شعاعية في التعابير السابقة بإجراء بعض التعديلات.

اقترحت الشبكة العصبونية ذات التراجع المعممة لإنجاز تراجع (انحدار) عام خطي أو غير خطي. وقد اقترح هذا النوع من الشبكات Specht عام 1991[218]. إن هذه الشبكات قادرة على إنجاز تحليل تراجعي مباشرةً من معطيات العينة. وباستعمال هذه الشبكة، ليس هناك افتراضات لازمة فيما يخص الشكل التابعي المرتبط بالمتحولات المستقلة وغير المستقلة كما في حالة التراجع الإحصائي.

تنجز الشبكة التقدير مباشرة من توزيع احتمال أساسي لنماذج الدخل باستعمال بعض تقنيات التقدير الموصوفة آنفاً.

من المفترض أن كل نموذج دخل **x** ينتمي إلى أحد التجمعات K حيث عدد النماذج المنتمية للتجمع j هو $k_j$.

قبل أن تصبح الشبكة مبنية، توضع نماذج التدريب في مجموعات لتشكيل تجمعات معروفة. يمكن أن تنجز عملية تشكيل التجمعات باستعمال أي طريقة من الطرائق العديدة مثل متوسطات K (K-means)، أو الجوار الأقرب، أو استعمال الشبكات العصبونية الصنعية كشبكة التعليم بالتكميم الشعاعي (LVQ)، أو خريطة الملامح ذاتية التنظيم (SOFM) لـ Kohonen (سنناقش ذلك في الفصل القادم).

إذا كان عدد نماذج التدريب مقبولاً (ليس ضخماً جداً)، فيمكن أن يعمل كل نموذج كأنموذج؛ وهذا يعني أن كل تجمع يحوي نموذجاً واحداً فقط (في هذه الحالة، $k_j=1$ لكل j). أيضاً ستكون أشعة نموذج الدخل قياسية أو معيارية من أجل إنجاز أفضل.

بعد معرفة عدد التجمعات والمراكز المتوسطة للتجمعات والنماذج المعيارية يمكن تصميم الشبكة وتدريبها. صممت بنية الشبكة وخوارزمية تدريبها لتسهل حسابات التحليل التراجعي الموصوفة آنفاً. لاحظ أن المعادلة (13.12) تطبق على تجمعات بنماذج مفردة فقط.

لتنفيذ التراجع في حالة التجمع العام، يجب أن يعدل الإجراء السابق. وفي هذه الحالة يجب أن يقدر $\widehat{z}$ بواسطة:

$$\widehat{z} = \frac{\sum_{i=1}^{P} A_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{P} B_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \qquad (16.12)$$

حيث

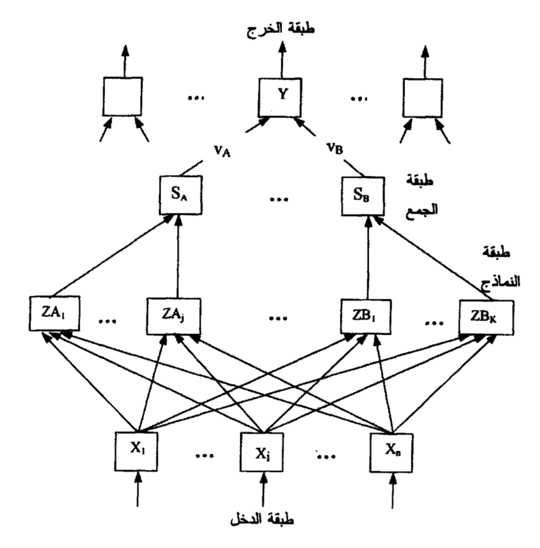
$$A_i \equiv A_i(k) = A_i(k-1) + z_j$$
$$B_i \equiv B_i(k) = B_i(k-1) + 1$$

هذه الكميات هي قيم الثوابت للتجمع $i$ بعد k مراقبة.

بنية الشبكة العصبونية ذات التراجع المعممة مشابهة لبنية الشبكة العصبونية الاحتمالية، فهي تتألف من طبقة دخل متبوعة بثلاث طبقات حسابية: طبقة النماذج وطبقة الجمع وطبقة الخرج، كما هو موضح في البنية الأساسية في الشكل (3.12). قارن هذه البنية مع بنية الشبكة العصبونية الاحتمالية في الشكل (1.12).

استعملت طبقة الدخل لتوزيع نماذج الدخل على طبقة النماذج التالية. تضم طبقة النماذج K وحدة، وقد خُصِّصت وحدة واحدة لكل نموذج أو وحدة واحدة لكل تجمع نموذج. هذه الطبقة متصلة كلياً مع طبقة الدخل من خلال أوزان معدلة. توضع قيم الأوزان مساوية لنماذج الأنموذج أو لقيم مراكز التجمع (قيم المراكز المتوسطة). وهكذا، فإن الأوزان على الوصلات من طبقة الدخل إلى الوحدة رقم $i$ في طبقة النماذج لها قيم مساوية لشعاع المراكز المتوسطة للتجمع (ذي البعد n).

عندما يقدم نموذج الدخل x إلى طبقة الدخل، تحسب وحدات طبقة النماذج المسافة بين شعاع وزنها وشعاع الدخل. بعدئذ، تحول هذه المسافة بواسطة تابع تفعيل الوحدة.




**الشكل3.12**: بنية الشبكة ذات التراجع المعممة

إن توابع تفعيل وحدات النماذج هي توابع أسية(لغوص) أو توابع مشابهة، حيث سيكون وسيط المسافة بين **x** وشعاع الوزن $\mathbf{w_j}$. مثلاً، إذا استعمل مجموع مربعات مركبات الفروق كمسافة مترية، فإن خرج الوحدة رقم $j$ في طبقة النماذج يعطى بالعلاقة:

$$f(\mathbf{x}, \mathbf{w}_j) = \exp\left[\frac{\sum_{i=1}^{n}(\mathbf{x}_i - \mathbf{w}_{ij})^2}{2\sigma^2}\right] \qquad (17.12)$$

حيث $\sigma$ ثابت النعومة.

يمكن أن تستعمل قياسات أخرى للمسافة، إضافة إلى المسافة الإقليدية المربعة، إضافة إلى

أنواع أخرى من توابع التفعيل. يوصل خرج طبقة النماذج وصلاً كاملاً مع وحدات طبقة الجمع بناءً على أوزان معدلة.

تظهر في الشكل (3.12) فقط وحدتا جمع ووحدة خرج وذلك بهدف تبسيط الشكل. على أية حال، يمكن إضافة وحدات جمع وخرج إضافية كما سيوصف فيما يلي.

طبقة الجمع مركبة من نوعين من العصبونات، نوع A ونوع B. إذا كان لطبقة الخرج أكثر من وحدة واحدة، لنفترض m وحدة، فسيكون هناك m نوع A وأيضاً m نوع B من الوحدات في طبقة الجمع.

تعالج مداخل وحدات طبقة الجمع بواسطة إنجاز جداء نقطي شعاعي (سلّمي) بين شعاع الدخل من طبقة النماذج وقيم شعاع وزن وصلاتها.

تستعمل العصبونات من النوع A لتمثيل مخارج التراجع المنشودة للنماذج. تساوي قيمة الأوزان على وصلات هذه الوحدات لمجموع العينات $z_i$ المرافقة للتجمع $\mathbf{x}_i$. وهذا يعطي خرجاً من هذه الوحدات مساوٍ لـــ $\hat{z}f(\mathbf{x})K$ ، حيث $\hat{z}$ هو تقدير متوسط شرطي لــــ Z مع $\mathbf{x}$ معطى ($E[z|\mathbf{x}]$)؛ أي تراجع z على $\mathbf{x}$، و$f(\mathbf{x})$ هو تابع كثافة $\mathbf{x}$، وK ثابت محدد بواسطة نافذة بارزن المستعملة (K ليست معطيات معتمدة على غيرها، ولا تحتاج إلى حساب).

من ناحية أخرى، يساوي كلُّ وزنٍ من أوزان الوحدة نوع B عددَ النماذج في التجمع. تحسب هذه الوحدات الكميات $f(\mathbf{x})K$ بواسطة إنجاز جداء نقطي على مخارج وحدات النماذج والأوزان الموافقة.

تُنجِز طبقةُ الخرج عمليةَ تقسيم على خرج وحدتـــي طبقة الجمع لإعطاء تقدير $\hat{z}$ بتراجع z على $\mathbf{x}$.

يحدد وسيط النعومة σ حد سطح القرار (الفصل) ويجب أن يعين تجريبياً. يؤدي استخدام قيم صغيرة لــ σ إلى أسطح شوكية ضيقة شديدة الانحدار تلبس جيداً (تنسجم) قرب نقاط العينة فقط. وفي حالة قيم كبيرة لــ σ سنحصل على أسطح ناعمة ومنبسطة لهضاب منحدرة ببطء. يتطلب التعميم الجيد قيماً بين هاتين القيمتين الحديتين لــ σ. غالباً تعطي قيم σ المحصورة ضمن المجال [2.6] إنجازاً جيداً لتطبيقات عديدة (Specht عام 1991

([218]

عملية تدريب الشبكة العصبونية ذات التراجع المعممة سريعة ومباشرة، حيث يلزم مرور واحد فقط على مجموعة التدريب. ليس هناك حاجة للحسابات التكرارية كما في حالة خوارزمية الانتشار الخلفي أو في خوارزميات تدريب الشبكات الأخرى لأن الشبكة تحسب التقديرات لـــ $\hat{z}$ (المعادلة (4.12)) مباشرة من العينات.

في الحقيقة، تبدأ الشبكة بتنفيذ التراجع بعد تقديم عينة تدريب مفردة إلى الدخل، وعندما يستمر تقديم العينات إلى دخلها أكثر فأكثر يستمر تحسن إنجاز الشبكة. على أية حال، يعتبر عبء العمليات الحسابية على الشبكة ثقيلاً نسبياً.

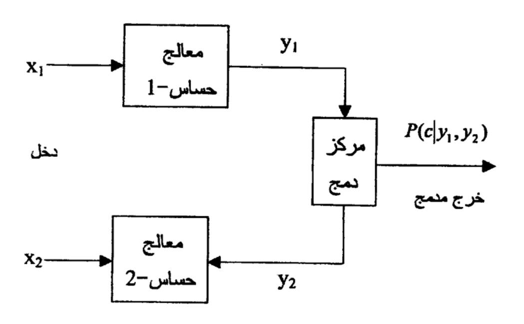
هناك منافع أخرى للشبكة العصبونية ذات التراجع المعممة تتمثل في مقدرتها على العمل بمعطيات مشتتة وفي أوساط الزمن الحقيقي وذلك لأن سطح التراجع يعرف في أي مكان لحظياً.

## 5.12 تطبيقات الشبكات المبنية احتمالياً

### 1.5.12 تصنيف علامات الاهتزاز لمطحنة تصنيع الفولاذ

أخذ أهم تطبيقات الشبكة العصبونية الاحتمالية هو تصنيف علامات (مؤشرات) الاهتزازات المجموعة من مطحنة تصنيع الفولاذ، ومراقبة الاهتزاز وتحليل المعطيات من سجلات "الانسياب الطبقي" في صحيفة مطحنة تصنيع الفولاذ التـــي تمكن من كشف الأخطاء المهددة بالوقوع، ومن ثمَ أخذ قياسات تصحيحية سلفاً.

في هذا التطبيق المنفذ من قبل Loskiewicz-Buczak وUhrig عام 1993[219]، جمعت المعطيات من حساسات موضوعة في تسعة أماكن على تسع آلات. جمعت الإشارات المستقاة من الآلات المختلفة تبعاً للحساس وأماكن الخطأ. جرى توليد طيف خرج كل حساس باستعمال تقنيات تحويل فورييه السريع FFT، وقد جمعت 150 نقطة معطيات وخزنت في قاعدة المعطيات. تحوي مجموعة المعطيات علامات (مؤشرات) من 49 آلة، يوافق كل منها تشخيص نوع من الأخطاء.




**الشكل 4.12**: دمج وتصنيف مدخل حساسين اثنين

استُعمل دمج إشارات عدة حساسات لمسائل تصنيف الخطأ. يبين الشكل (4.12) عملية الدمج والتصنيف لإشارات حساسين اثنين.

يعطي كل حساس قراراً على أساس المعطيات من أحد حساسات الدخل. يستعمل بعدئذ مركز دمج الصفوف المتداخلة لجعل قرار التصنيف نهائياً. يحول شعاعا الدخل $\mathbf{x}_1$ و$\mathbf{x}_2$ أولاً بواسطة معالجات الحساسات $S_1$ و$S_2$. تكون بعدئذ المخارج المحولة $y_1$ و$y_2$ مدمجة وينفذ تقدير الصف بواسطة الشبكة العصبونية الاحتمالية وفقاً للعلاقة:

$$P(c|\mathbf{x}) \propto P(c|\mathbf{x}_1)P(c|\mathbf{x}_2)\left[\frac{P(c|\mathbf{y}_1,\mathbf{y}_2)}{P(c|\mathbf{y}_1)P(c|\mathbf{y}_2)}\right]$$

حيث $P(c|\mathbf{x}_1)$ و$P(c|\mathbf{x}_2)$ هما احتمالات التصنيف المستقلة، والعبارة الكسرية في الأقواس هي عامل التصحيح عندما توجد الارتباطات بين $\mathbf{x}_1$ و$\mathbf{x}_2$ (هذا الحد له قيمة تساوي $1/P(c)$ في حالة الاستقلال). ويمثل البُسط في علاقة عامل التصحيح الكسرية تابع قرار بايز الصحيح للحساسات التسعة.

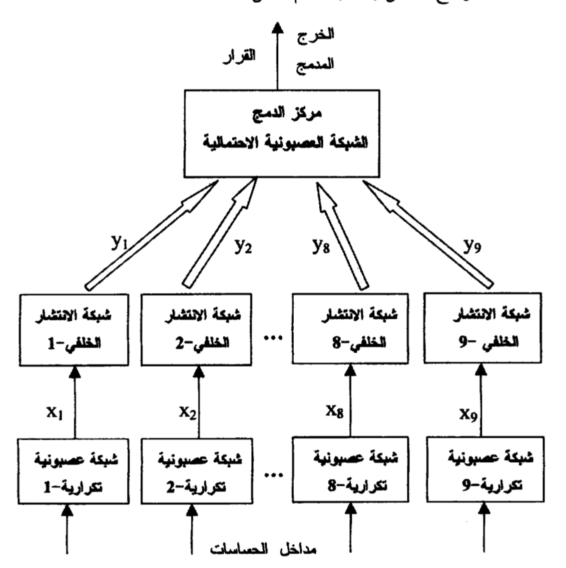
صُنِّفت مؤشرات الاهتزاز بثلاثة أطوار:

يشمل الطور I: استخلاص الملامح المناسبة من طيف الاهتزاز لكل حساس بوجه منفصل،

ومن ثم ضغط وتحويل الطيف باستعمال شبكات عصبونية بتنفيذ متكرر.

يشمل الطور II: تصنيف عملية ضغط المعطيات باستعمال شبكة أمامية التغذية بخوارزمية الانتشار الخلفي، عادة واحدة لكل حساس( الفصل السادس).

ينجز الطور III: دمج معلومات القرار من المصنفات الفردية باستعمال الشبكة العصبونية الاحتمالية. يوضح الشكل (5.12) النظام ككل.




**الشكل 5.12**: البنية الكاملة لنظام التصنيف

تأخذ كل من شبكات التنفيذ المتكرر 150 نقطة ـــ مؤشر كدخل معطيات من حساس

معطى وتضغطها حتـــى 50 نقطة. تستعمل المعلومات المضغوطة كدخل لشبكة الانتشار الخلفي، حيث سيكون هناك شبكة واحدة لكل حساس.

تعطي شبكات الانتشار الخلفي قراراً على درجة المؤشر المعطى في انتمائه إلى كل من الصفوف. أخيراً، ستكون قرارات شبكات الانتشار الخلفي دخلاً لمركز الدمج والتصنيف النهائي.

### 2.5.12 تصنيف مخططات القلب Classification of Electrocardiogram

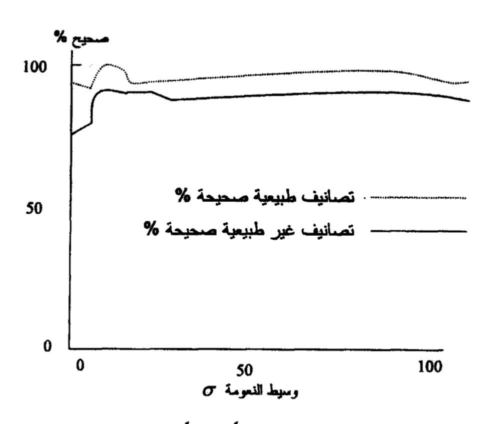
يتطلب تحليل مخططات القلب خبرات مدربة للتمييز بين حالات القلب الطبيعي وغير الطبيعي. يتطلب التحليل فحصاً دقيقاً لنماذج المخططات البيانية من عدة حساسات، وذلك لمعرفة فعالية قلب المريض خلال مدة قصيرة.

في دراسة مقترحة من قبل Specht عام 1967 [220]، تم تدريب شبكة عصبونية احتمالية على 249 نموذج لمريض بشروط معروفة. دربت الشبكة لتصنيف المرضى إلى أحد صفين: طبيعي وغير طبيعي.

يتألف نموذج الدخل من شعاع ذي 46 مركبة. في الاختبار، استعمل 63 نموذجاً إضافياً. يوضح الشكل (6.12) النسبة المئوية التقريبية للتصانيف الصحيحة كتابع لوسيط النعومة. يتضح من الشكل أن $\sigma$ هو الوسيط الوحيد الذي يجب أن يحدد تجريبياً، ويمكن عادة إيجاد قيمة مناسبة من بضع اختبارات. أُجري تقريب واحد بمقارنة مستويات الدقة الناتجة لقيم مختلفة لـــ $\sigma$ عندما استعملت عينات مفردة للتدريب ومن ثم للاختبار. يلزم فقط مجال محدد من القيم للاختبار. مثلاً، مـــن الشكل (6.12) نـــرى أن دقة التشخيص الأعظمية تحققت لقيم $\sigma$ بين 4 و6.

لقد تبين أن إنجاز الشبكة لا يتأثر كثيراً باختيار $\sigma$، ومن ثم فإن إيجاد قيمة مناسبة تجريبياً ليس عملية صعبة.

بالفعل يمكن إثبات أن أي قيمة في المجال من 3 حتـــى 10 تحقق نتائج جيدة، وأيضاً، تبقى القيم خارج هذا المجال أفضل من التصانيف المختارة عشوائياً.




**الشكل 6.12**: النسبة المئوية للعينات المصنفة تصنيفاً صحيحاً بواسطة الشبكة العصبونية الاحتمالية

### 3.5.12 نمذجة ديناميكيات الطائرات Modeling Aircraft Dynamics

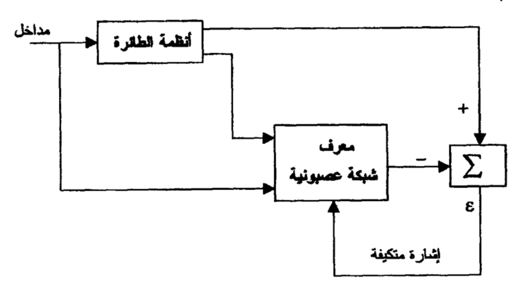
يمكن أن تستخدم الشبكات العصبونية ذات التراجع المعممة في تطبيقات عديدة منها التنبؤ، ونمذجة نظام غير معروف، وكذلك في التحكم، ومسائل التطبيق العام. سنصف في هذه الفقرة نمذجة الديناميكيات غير الخطية للطائرات المقاتلة. إن الديناميكيات الفضائية للطائرة المقاتلة هي نموذجياً غير خطية أبداً في الطبيعة.

تحدث عدم الخطية هذه نتيجة لارتباطات الطاقة الكامنة والطاقة الحركية، ولعدم خطيات ديناميكيات الفضاء، وللقيم المحددة لمعدل التحكم بالانحراف.

في دراسة قدمها H. M. Youssef عام 1993[136]، نفذت المحاكاة لتوضيح كيف تستطيع بنـــى الشبكة العصبونية الصنعية المختلفة نمذجة ديناميكيات الطائرة المقاتلة الأمريكية F-16 من أجل مناورتين جويتين غير خطيتين أبداً: زوايا منخفضة للانقضاض والهبوط العميق.

أعطيت الشبكة خمسة متحولات متعلقة بالزمن في لحظات زمنية متقطعة k، ومتحولي

استجابة متنبأ بهما.




**الشكل 7.12**: نموذج شبكة عصبونية لديناميكيات طائرة مقاتلة F-16

يَستعمل النظام مُعرِّفَ (Identifier) شبكة عصبونية تتعلم نموذج ديناميكيات الطائرة كما هو موضح في الشكل (7.12). كانت متحولات دخل الشبكة هي أمر التحكم بالانحراف $\delta_c(k)$، وزوايا الهجوم عند $\alpha(k)$ و $\alpha(k-1)$، ومعدلات درجة الانحدار عند $q(k)$ و $q(k-1)$.

تألف شعاع الخرج من التنبؤ بزوايا الهجوم $\bar{\alpha}(k+1)$ ومعدل الانحدار $\bar{q}(k+1)$ عند اللحظة الزمنية k +1. تعطى أخطاء التقريب لزوايا الهجوم ومعدل الانحدار بالكميات التالية:

$$E_a = \sqrt{\sum_k \left(\alpha(k) - \bar{\alpha}(k)\right)^2}$$

$$E_q = \sqrt{\sum_k \left(q(k) - \bar{q}(k)\right)^2}$$

في هذه الدراسة، أجريت دراسة مقارنة لأداء ست من الشبكات العصبونية: شبكة إنجاز التراجع الخطي (Linear Regression)، وشبكة كثير الحدود من المرتبة الثالثة (third-order polynomial) وشبكة متعددة الطبقات أمامية التغذية (بطبقتين مخفيتين بالانتشار الخلفي، وشبكة عنصر تحكم بوصلات مفصلية لنموذج مخيخي CMAC؛ (Cerebellar Model Articulation Controller network)، وشبكة تابع الأساس

الشعاعـــي RBF (Radial basis function)، وأخيـــراً شبكـــة التراجع المعممة Generalized Regression. بنيت مقارنة الأداء لهذه الشبكات على أساس سرعة التعليم، ودقة النمذجة، ومرونة الشبكة، وتعقيد البنية.

في هذه الدراسة، تطلبت الشبكة المتعددة الطبقات الأمامية التغذية بخوارزمية الانتشار الخلفي 50 عقدة طبقة مخفية وأكثر من 50 دوراً حتـــى تحقق التقارب، أما شبكة تابع الأساس الشعاعي فقد استعملت 18 عقدة أساس لتمييز المتحولات، وهذه الشبكة والشبكات الأخرى تقاربت بعد بضعة أدوار، لكن الشبكة العصبونية ذات التراجع المعممة دربت خلال مرور واحد عبر مجموعة التدريب.

الشبكة العصبونية ذات التراجع المعممة GRNN ولّدت بـــ 1600 عقدة من أجل التجربة الثانية (200 من أجل التجربة الأولى). كان مجال قيم $\sigma$ بيـــن [0.1-5] مقنعاً. ومع أن إنجاز كل الشبكات لهذه المسألة كان مقنعاً ماعدا شبكة التراجع الخطي، فإن إنجاز شبكة التراجع المعممة كان هو الأفضل. وأصبح هذا واضحاً من خلال مقارنة الأخطاء لقيم $E_a$ و $E_q$. يعطي الجدول 1.12 ملخصاً لقيم هذه الأخطاء.

الجدول 1.12 تقريب الأخطاء $E_a$ و $E_q$ لمناورتين للطائرة المقاتلة

| نوع الشبكة | الحالة -1 | الحالة -1 | الحالة -2 | الحالة -2 |
|---|---|---|---|---|
| | $E_a$ | $E_q$ | $E_a$ | $E_q$ |
| التراجع الخطية | 2.4571 | 1.2220 | 10.3168 | 6.4562 |
| كثير الحدود | 2.2207 | 1.1490 | 7.4597 | 4.3873 |
| متعددة الطبقات | 2.8777 | 2.7712 | 3.7296 | 2.7893 |
| الأساسي الشعاعي | 2.145 | 1.3625 | 4.1159 | 3.2475 |
| CMAC | 2.3595 | 4.9421 | 4.3139 | 3.4102 |
| التراجع المعممة | 2.0910 | 1.2176 | 2.8319 | 2.7865 |

# خريطة الملامح الذاتية التنظيم والتكميم الشعاعي
# Self-organizing feature Map and Vector Quantization

هذا الفصل الأول أحد فصلين اثنين سيُستعمل فيهما التعليم بدون معلم. يمكن أن يقسم التعليم بدون معلم إلى صنفين اثنين: التعليم التنافسي والتعليم غير التنافسي (Hebbian). سنعطي أولاً وصفاً للصنفين، ثم سنقوم بمقارنة الاختلاف بينهما، وسنتبع ذلك ببعض التطبيقات الخاصة.

سنبحث بوجه خاص في الشبكات القادرة على تعلم نوع فعال للتكميم الشعاعي (Vector Quantization) VQ.

التكميم هو عملية تحويل متحول بقيمة تماثلية (analog) أو مستمرة إلى متحول متقطع. تتعلم شبكات التكميم الشعاعي VQ تكميمَ وترميزَ نماذج الدخل من وسط ما، كما في الشبكات البيولوجية. سنهتم بتغيرين أساسيين لشبكة التكميم الشعاعي وسنشير إليهما بشبكة التكميم الشعاعي 2(VQ2) وشبكة التكميم الشعاعي 3 (VQ3). سينظر إلى شبكة التكميم الشعاعي3 كخريطة ملامح ذاتية التنظيم (Self-organizing feature map) SOFM، وهي تعميم لعملية التكميم الشعاعي.

سندرس هذه البنية بالتفصيل وذلك لمقدرتها المذهلة على التطبيق. ولن ننسى بالطبع أن نعرّج على بعض أنواع الشبكات المبنية على التنافس، ومن ثم عرض لبعض تطبيقات التكميم الشعاعي، وتطبيقات خريطة الملامح الذاتية التنظيم.

وأخيراً سنناقش البنـــى المختلفة المبنية على التعليم بدون معلم وعائلة نظرية الطنين المتكيف أو شبكات الطنين المتكيف ART (Adaptive Resonance Theory) (في الفصل القادم).

## 1.13 تمهيد

الشبكات المدروسة في هذا الفصل هي ذاتية التنظيم (self-organizing)؛ أي إنها تتعلم بدون إشراف من معلم.

خلال عملية التعليم، تقدم متتالية نماذج الدخل x إلى الشبكة، حيث تولد النماذج بواسطة توزيع احتمالي ما ρ(x) غير معروف عادةً. عندما يقدم النموذج إلى الشبكة تستجيب لحساب تفعيلات الخرج، ولكن لن يكون هناك تغذية عكسية مباشرة معطاة إلى الشبكة لتصحيح الاستجابة للدخل. بالفعل لن يكون هناك جواب صحيح! وكذلك لن يكون هناك مؤشر أو دليل عن كون الخرج صحيحاً أو خطأً !!. يجب أن تتعلم الشبكة بطريقة أو بأخرى اكتشاف أو استغلال أي بنية موجودة بين نماذج أو أمثلة الدخل. ما هي أنواع البنـــى التـــي يمكن أن تكتشفها الشبكة من مجموعة أمثلة الدخل؟.الجواب عن هذا السؤال يعتمد بالطبع، على المنبع ρ(x) وعلى إجراءات التعليم المستخدمة في الشبكة.

عموماً، يمكن أن توجد أنواع البنـــى التالية:

1. مجموعات (groupings) أو تجمعات (clusters) النماذج المرتبطة بعضها ببعض بشدة.
2. عدد تكرارات حدوث مجموعات النماذج.
3. المراتب النسبية (الطول) فيما بين المداخل الشعاعية.
4. الارتباطات فيما بين النماذج (وبوجه خاص، يمكن أن تكتشف الشبكة أي متحولات مركبة شعاعية لها أعظم تغيرية؛ شكل لتحليل المركبة الأساسية).
5. التطبيقات (mapping) التـــي تحول نماذج الدخل إلى فراغ ببعد أخفض ؛ نوع من الترميز (coding) المبنـــي للمداخل.
6. تطبيق الململح: تحويل الجملة المولدة (manifold) للدخل إلى أخرى ببعد مختلف أثناء عملية المحافظة على الطبولوجية (بنية خطية، أو مستطيلة، أو سداسية، تفرض ما بين الوحدات).

تقع طرق التعليم بدون معلم التـــي سندرسها ضمن إحدى فئتين: إجراءات تنافسية وغير تنافسية. أما الشبكة فقد تكون بطبقة مفردة أو بعدة طبقات.

سينصب اهتمامنا على النوع التنافسي وذلك لشعبيته الواسعة. هناك أصناف عديدة

للشبكات غير التنافسية درست من قبل باحثين عديدين مثل Linsker عام 1988[6] و Oja عام 1982[221]. استعملت جميع هذه الشبكات طرقاً مبنية على نموذج ما من تعليم Hebb أو Hebb المعدل.

## 2.13 شبكات التعليم بدون معلم غير التنافسية

## Unsupervised Noncompetitive Learning Networks

بنيت كل الشبكات التـي سنستعرضها هنا على التعليم بدون معلم غير التنافسي، أي على نموذج ما من تعليم Hebb. تسمح نماذج Hebb بالتعليم الفعال بدون معلم لأن الوحدة التـي هي أكثر استجابة لدخل معطى يسمح لها بأن تخضع لتعليم أكثر من الوحدات التي هي أقل استجابة.

إذا كانت هناك مداخل متشابهة تعالجها تكرارياً بعض الوحدات، فسيكون لتلك الوحدات استجابة أكثر لهذا التجمع من النماذج المتشابهة، تاركة الوحدات الأخرى تكتشف تجمعات نماذج مختلفة. خلال هذه العملية، تتعلم الوحدات الاستجابة لتجمعات مختلفة معاً دون الحاجة لمعرفة أن هذا النموذج عضو في التجمع أم لا. تذكر أن المعادلة المبسطة لتعليم Hebb هي:

$$\Delta w_i = \alpha x_i y \qquad (1.13)$$

حيث $w_i$ الوزن على الوصلة من الدخل إلى الوحدة $i$، و $\alpha$ معدل التعليم، و $x_i$ دخل الوحدة $i$، و $y$ استجابة الوحدة $i$ لهذا الدخل.

لكي يكون هذا النوع من التعليم مفيداً يجب أن يخضع لبعض الشروط والقيود أو أن يعدل، وإلا فإن الأوزان يمكن أن تنمو بدون حد والتعليم قد لا يستقر مطلقاً. للحدِّ من هذه المشكلة وتهذيبها، اقترحت أشكال معدلة (1.13) مثلاً، يمكن أن تضم قاعدة تحديث الأوزان عامل إعادة المعايرة، أو يمكن أن تقيد الأوزان بوضعها ضمن قيم محددة أثناء تثبيتها عند تلك القيم، أو إضافة حد النسيان للتخفيض والحد من نمو الأوزان. تأخذ التعديلات باستعمال حد النسيان أو الاضمحلال المعادلات التالية:

$$\Delta w_i = \alpha (x_i y - y w_i) \qquad (2.13)$$

$$\Delta w_i = \alpha (x_i y - y^2 w_i) \qquad (3.13)$$

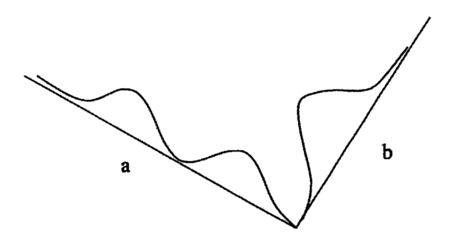
## تحليل المركبة الأساسية:

اقترحت قاعدة مبنية على المعادلة (3.13) من قبل Oja عام 1982[221]، حيث أثبت أن الأوزان **w** تتقارب فعلياً إلى الطول الواحدي بدون إعادة المعايرة. وهذا يعنـــي، أن الأوزان عندما تنضج تميل إلى تحقيق العلاقة $\sum_i w_i^2 = 1$ .

وبعد تقدم التعليم لبعض الوقت، تقع الأوزان الناضجة في اتجاه الشعاع الخاص الأعظمي لمصفوفة ارتباط نماذج الدخل، $\mathbf{C} = \mathbf{E}[\mathbf{x}\,\mathbf{x}^T]$. وهكذا، عندما تصبح الأوزان مستقرة، تتعلم الوحدة فعلياً إنجاز تحليل المركبة الأساسية **PCA (Principal Component Analysis)** على المداخل، وهي الطريقة الشائعة لتحليل المعطيات الإحصائية المستعملة لاستنباط الملامح (راجع الفصل الثالث).

يتطور شعاع الوزن **w** بحيث تكون نماذج الدخل مسقطة على المحور الموازي لـــ **w**، وفي تلك الحالة، سيكون لها تباين أعظمي. إن إسقاط **x** على المحور الموازي لـــ **w** عندما $\sum_i w_i^2 = 1$ ، يكون في الحقيقة مساوياً لخرج الوحدة $y = \sum_i x_i \cdot w_i$ ، وتباين التوزيع المسقط هو نفسه تباين y.

في تنفيذ تحليل المركبة الأساسية، تعلمت الوحدة أن تختار تلك المركبات المتغيرة لأشعة الدخل التـــي لها التأثير الأعظم أو التغيرية العظمى على الخرج **y**. الفائدة من هذه الطريقة موضحة في الشكل (1.13) في حالة شعاع دخل بمتحولين فقط.




**الشكل1.13:** تحليل المركبة الأساسية لمتحولين

نلاحظ من الشكل أنه إذا كانت مجموعة المعطيات مشاهدة من المحور المشار إليه بـــa، فإن التوزيع يبدو بوضوح ثنائيَّ النمط مع تباين كبير بشكل معاكس للمسقط على المحور b الذي يكون بنمط مفرد فقط ملاحظ وتوزيع بتباين أصغر. وهكذا، في معطيات التحويل الأصلية من الشكل الأخير إلى الشكل المشكل، تساعد تحويلات تحليل المركبة الأساسية على تعرية البنية في المعطيات التـــي قد تكون مخفية. هناك نقطة أخرى لتحليل المركبة الأساسية يمكن أن تكون مفيدة جداً، وهي تقليل بعد فراغ الملمح. نستطيع باختيار مركبات الشعاع التـــي تأخذ بعين الاعتبار التغيرية العظمى فقط وبإمكانية حذف المتحولات غير المساهمة، تقليل بعد فراغ الدخل.

بالطبع، من أجل شبكة بوحدة مفردة، توجد فقط أول مركبة أساسية. لإنجاز تحليل مفيد لمركبة أساسية متعددة المتحولات يجب أن يكون للشبكة مخارج متعددة. فيما يلي سنصف حالات أخرى من الشبكات التـــي تتعلم إنجاز تحليل المركبة الأساسية، بما في ذلك تعريف المركبة المتعددة المتحولات.

طور الباحثان Oja عام 1989[222] وSanger عام 1989[223] الشبكات الخطية الوحيدة الطبقة بمخارج متعددة. بالإضافة إلى ذلك فقد طورت شبكة متعددة الطبقات خطية ودرست من قبل Linsker عام 1988[6] ستشرح لاحقاً.

أيضاً درست قاعدة تعليم Hebb معدلة أخرى مرتبطة بالقواعد المعطاة بالمعادلات (2.13) و(3.13) من قبل Yuille وزملائه عام 1989[224]. فقد استَعملت قاعدتهم حدَّ الاضمحلال المتناسب مع نظيم الأوزان **w** المربع عوضاً عن الخرج **y**. هذه القاعدة هي كما يلي:

$$\Delta w_i = \alpha(x_i y - w_i \|w\|^2) \qquad (4.13)$$

لقد أثبت أن شعاع الوزن لهذه القاعدة يتقارب أيضاً إلى نهاية محددة وفي نفس الاتجاه كالشعاع الخاص الأعظمي لمصفوفة الارتباط **C**. تكون القيم النهائية المحددة للأوزان في هذه الحالة مساوية للقيمة الخاصة الأعظمية.

كما ذكر من قبل، درست شبكات التغذية الأمامية بمخارج متعددة باستعمال نماذج قاعدة Hebb المعدلة. كانت توابع التفعيل لهذه الشبكات خطية بخرج $y_i$ معطى كما يلي:

$$y_j = \sum_{i=1}^{n} x_i w_{ij} \quad , \quad j = 1,2,\cdots m \qquad (5.13)$$

هناك قاعدتا تعليم متشابهتان تَستعملان حدود اضمحلال أيضاً، تتألف من جداءات مخارج الوحدة والأوزان، وهذا يشبه إلى حد ما قاعدة تعليم الانتشار الخلفي. إحداهما معطاة بالمعادلة (6.13) اقترحها Sanger عام 1989[223]:

$$\Delta w_{ij} = \alpha y_i (x_j - \sum_{k=1}^{i} y_k w_{kj}) \qquad (6.13)$$

حيث يلاحظ من هذه المعادلة أن حد المجموع العلوي هو تابع للأوزان الموجودة للتحديث.

والقاعدة الثانية هي نفسها القاعدة السابقة ماعدا أن حد المجموع العلوي في هذه القاعدة يساوي n، وهو العدد الكلي للمداخل( Oja عام 1989[222]):

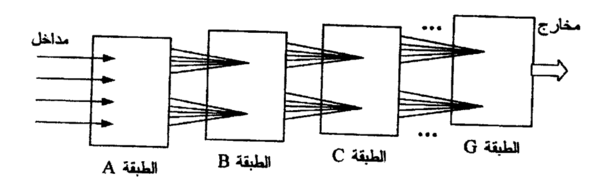
$$\Delta w_{ij} = \alpha y_i (x_j - \sum_{k=1}^{n} y_k w_{kj}) \qquad (7.13)$$

وسَّع Sanger عام 1989[223] نتائجه إلى حالة غير خطية مع توابع تفعيل sigmoid ألزمت بالصفر (عند قيمة العتبة). وقد استُعملت هذه الشبكات في مسائل التكميم الشعاعي وضغط المعطيات.

## 3.13 الشبكات المتعددة الطبقات بدون معلم

### Unsupervised Multilayer networks

اقترحت بنية مختلفة نوعاً ما من قبل Linsker عام 1988[6]، الذي درس شبكات متعددة الطبقات أمامية التغذية باستعمال نموذج من قاعدة Hebb. انصب اهتمام Linsker على تحديد إمكانية تعلُّم الوحدات الذاتية التكيف الخطية تطوير مقدرات تحليل الملمح المفيد بأسلوب مشابه لتلك المكتشفة في القشرة البصرية. مثلاً، من المعروف أن ملامح المستوى المنخفض البسيطة، مثل زوايا الوصل والاتجاهات المتغايرة، تعالج في المراحل المبكرة من النظام البصري، على حين تعالج ملامح المستويات العليا مثل الوصلات المتعددة والأشياء الكاملة في المراحل العليا. ركبت الشبكات المستعملة لنمذجة هذا النوع من المهام من طبقات متعددة بوحدات ثنائية البعد كما هو موضح في الشكل (2.13).




**الشكل 2.13:** شبكة Linsker أمامية التغذية بطبقات ثنائية البعد من الوحدات

جرى تسمية الطبقات بالأحرف A , B , C , D , E , F , G على الترتيب، حيث تستقبل الطبقة A المداخل الخارجية من الوسط المحيط وتعطي الطبقة G مخارج المستوى العالي. وتستقبل وحدة واحدة في أي من الطبقات المتتالية دخلاً من مجموعة أو من جوار من الوحدات في الطبقة السابقة. ولكلّ وحدة من الوحدات في الطبقات المتتالية عدةُ مئات من وصلات الدخل القادمة من وحدات الطبقة السابقة. وجميع هذه الوحدات لها توابع تفعيل خطية من الشكل:

$$y = \beta + \sum_i x_i w_i \qquad (8.13)$$

حيث $\beta$ ثابت، و $x_i$ ، $w_i$ هما مداخل وأوزان الوصلات من الطبقة السابقة على التتالي.

ومع أن خرج الشبكة المتعددة الطبقات بتوابع تفعيل خطية يساوي إلى تحويل جداء خطي للخرج بطبقة وحيدة (راجع الفقرة 1.5)، فقد سُمح للطبقات بأن تتطور خلال التدريب بأسلوب تتابعي لكل طبقة (طبقة تلو الأخرى) حتـــى تستقر الأوزان استقراراً كاملاً. أولاً، ستدرب الأوزان بين الطبقة A والطبقة B، وعندما تستقر هذه الأوزان يبدأ بتدريب الأوزان بين الطبقة B والطبقة C حتـــى تستقر، وهكذا.

ولتنفيذ المحاكيات، تستقبل الوحدة في الطبقة A نماذج دخل عشوائية من العالم البصري. أما المداخل العشوائية المقدمة إلى الطبقة A والمقدمة إلى مداخل وحدات الطبقة B فهي نفسها (مصفوفة التباين المتبادل متناسبة مع المصفوفة الواحدية). شكل قاعدة Hebb المستعملة لكل الوحدات معرفة بما يلي:

$$\Delta w_i = \alpha_1 x_i y + \alpha_2 x_i + \alpha_3 y + \alpha_4 \qquad (9.13)$$

حيث $\alpha_i$ عوامل كيفية مع وسيط معدل التعليم $\alpha_1 > 0$. في هذا النموذج، منعت الأوزان من النمو غير المحدود بواسطة تقليمها؛ أي بواسطة تحديد قيمها لتقع ضمن المجال $w_- \leq w \leq w_+$، حيث $w-$ و$w+$ هما قيم الحد الأدنـــى والأعلى للمجال على الترتيب.

من المفيد معرفة كيف تتطور الأوزان وفقاً للمعادلة (9.13) عندما يقدم عدد ضخم من نماذج الدخل إلى الشبكة. يمكن أن نوجد التوقع الرياضي أو معدل التغير $E[\Delta w_i]$ بتعويض قيمة y (المعادلة (8.13)) في (9.13) وإيجاد القيمة المتوقعة المشار إليها. ولما كانت $x_i$ موزعة توزيعاً متماثلاً بقيمة متوسطة $E[x_i] = \mu$، ولتكن $x_i = z_i + \mu$. فإن:

$$E(\Delta w_i) = \alpha_1 E\left[(z_i + \mu)\left(\beta + \sum_i (z_i + \mu) w_i\right)\right]$$

$$+ \alpha_3 E\left[\sum_i (z_i + \mu) w_i\right] + k \qquad (10.13)$$

$$= k_1 + \sum_j C_{ij} w_j + \frac{k_2}{n} \sum_j w_j$$

حيـــث $k$ و $k_1$ و $k_2$ حـــدود ثابتـــة أخـــذت كتـــراكيـــب لـــ $\alpha_1 - \alpha_4$، $\mu$، و $C_{ij} = E[(x_i - \mu)(x_j - \mu)]$ عناصر مصفوفة التباين المتبادل C لأشعة الدخل. سنرى عما قريب أن المصفوفة C تؤدي دوراً هاماً في تطوير الوحدات في كل طبقة من طبقات الشبكة.

عندما تعطى نماذج الدخل العشوائية للطبقة A، فإن قيم $C_{ij}$ لتفعيلات خرج الطبقة A ستكون ببساطة مساوية للواحد عندما يكون $i = j$ ومساوية للصفر في الحالات الأخرى. هذه القيم، مع مخارج وحدات الطبقة A تعين قيم الأوزان المتطورة على الوصلات من الطبقة A إلى الطبقة B.

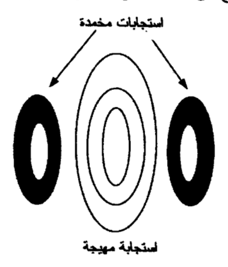
وبالمثل فإن قيم $C_{ij}$ لتفعيلات الطبقة B مع مخارج وحدات الطبقة B تعين قيم الأوزان المتطورة على الوصلات من الطبقة B إلى الطبقة C، وهكذا تستمر نفس الإجراءات في باقي الطبقات المتتالية.

أثبتت المحاكيات المنفذة في حالة قيم مختلفة للوسطاء في المعادلة (10.13) أن أوزان وحدات الطبقة B تشبع عند النقطة $w+$، وعند هذه النقطة تحسب الوحدات متوسط الفعالية المحلي من المنطقة التـــي تستقبل فيها المداخل من الطبقة A. تطور الوحدات المجاورة في

الطبقة B الفعالية الارتباطية. عندما تكون الوحدة B فعالة (on)، فإن الوحدات المجاورة لها تميل إلى أن تكون فعالة أيضاً. هذا الشكل من تنظيم الوحدة في B يحـرض شكـل "مركز محاط" من الفعالية في وحدات الطبقة C التـي تعمل كمرشح متحسس للتغاير الذي يكون أكثر استجابة للبقع البيضاء اللامعة الدائرية المركزة عند حقل استقبال الوحدة المحاطة بخلفية سوداء.

إن وحدات المركز المحاط التـي تستجيب بطريقة معاكسة تتطور أيضاً. وهذا يعنـي، أنها تستجيب استجابة أعظمية للبقعة السوداء المحاطة بخلفية بيضاء لامعة. تحدد $C_{ij}$ لوحدات الطبقة C سلوك الطبقة التالية التـي تبدي سلوكيات توجيه اختياري. وتستجيب وحدات الطبقة D استجابة أعظمية للملامح مثل الحواف (الزوايا) أو القضبان باتجاهات خاصة. يوضح الشكل (3.13) حقل الاستقبال لوحدة ما. في هذه الحالة، يكون توجيه الوحدة عمودياً، لكن بوجه عام، عندما تستعمل وصلات التغذية الأمامية فقط، يكون تطور التوجيه كيفياً نوعاً ما. من ناحية أخرى، إذا استعملت وصلات جانبية للوحدات داخل الطبقة، سيتطور التوجيه في اتجاهات أخرى أيضاً.

تثير نقاط الإضاءة في المستوي استجابات خرج متناسبة مع قيمة المحيط عند تلك النقطة. وتتطور الوحدات التـي لها أفضليات توجيه متشابهة لتشغل مناطق شريطية الشكل غير منتظمة أيضاً. وجد هذا النوع من الانتقائية في القشرة البصرية عند القطط والقردة.




**الشكل3.13**: حقل استقبال بحقول توجيه اختياري مستعملة لكشف الملمح

## 4.13 خواص الاستمثال Optimization Properties

بعد أن نظرنا في أوجه تحليل ملمح متكيف لهذه الشبكات، سنعود لمناقشة بعض الخواص الهامة الأخرى المقدمة من قبل قاعدة التعليم نوع Hebb المعرفة بالمعادلة (9.13). بوجه خاص، سنرى بأن أوزان الوحدات تتطور إفرادياً لكي تجعل التباين الإحصائي لتفعيلات مخارج هذه الوحدات أعظمياً. تحت شروط معينة، يكافئ هذا جعل معدل تمرير المعلومات من دخل الوحدة إلى الخرج أعظمياً. بكلمات أخرى، تختار عملية التعليم مجموعة من الأوزان التـــي تجعل المحافظة على الأوزان أعظمياً.

نستطيع تحليل سلوك الوحدة خلال عملية التكيف بمراقبة تفعيلات الخرج عند نضج الأوزان. لهذا، سنعرف التابع e المؤلف من حدين؛ الحد الأول متناسب مع تباين تفعيلات الخرج، والحد الآخر تابع لقيم أوزان الوحدة. بالإضافة إلى ذلك، نرغب أن يكون التابع e معرفاً بحيث يكون:

$$-\frac{\partial e}{\partial w_i} = \frac{\partial w_i}{\partial t}$$

لكل $i$، وبحيث يتناقص e على طول ممر الانحدار الشديد المحلي كلما تغيرت الأوزان. وهكذا، عندما تستقر الأوزان عند $dw_i/dt = 0$ ، فإن e عندئذ سيكون الأصغر المحلي.

من المعـــروف أن الأصغر المحلي هو فعلياً الأصغر الكلي القريب (Linsker عام 1988 [6]). يعطى التابع e المحقق للشروط السابقة بالعلاقة التالية:

$$\begin{aligned} e &= -\frac{1}{2}\left[E\left[(y-\mu)^2\right] + k_1\sum_i w_i + \frac{k_2}{2n}\left(\sum_i w_i\right)^2\right] \\ &= -\frac{1}{2}\left[\sum_i\sum_j C_{ij} w_i w_j + k_1\sum_i w_i + \frac{k_2}{2n}\left(\sum_i w_i\right)^2\right] \end{aligned} \tag{11.13}$$

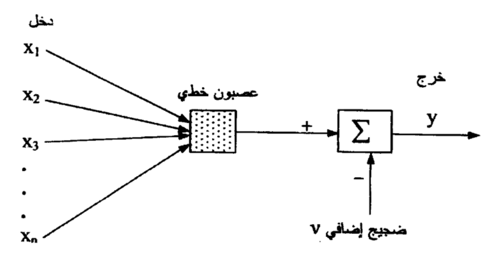
حيث E هو التوقع الرياضي، $\mu = E[y]$ ، مع $y = \sum_i x_i w_i$ ، و $C_{ij}$ هي حدود التباين المتبادل المعرّفة آنفاً. لاحظ أنه من أجل مجموعة من الأوزان، يكون e أصغرياً عندما يكون تباين المخارج أعظمياً. **وهكذا، تنضج الوحدة باختيار مجموعة من الأوزان التـــي تجعل**

**تباين تفعيلات الخرج أعظمياً لمجموعة نماذج الدخل.**

إن اختيار مجموعة الأوزان التـــي تجعل تباين تفعيلات خرج الوحدة أعظمياً، يعنـــي أن الوحدة ستتطور بطريقة تحافظ فيها على مقدار عالٍ من تمرير المعلومات من نماذج الدخل إلى تفعيلات الخرج. وبالفعل، يمكن أن يثبت ذلك لتكون الحالة هذه لتوزيعات نماذج دخل معينة.

مثلاً، افترض أن دخل وحدة قد أفسد ببعض ضجيج المعالجة $v$، حيث $v$ له توزيع طبيعي بمتوسط صفري وتباين $s_v$ كما هو موضح في الشكل (4.13). وافترض أن الخرج y أيضاً له توزيع طبيعي مع تباين $s_y$. وليكن الضجيج ومركبات الدخل $x_i$ غير مرتبطة بحيث $E(x_i v) = 0$ لكل $i$. بسبب كون الضجيج إضافياً، فإن خرج الوحدة سيكون له توزيع طبيعي أيضاً بمتحول عشوائيy معطى بالعلاقة التالية:

$$y = \sum_i x_i w_i + v \qquad (12.13)$$




**الشكل 4.13**: وحدة مشوشة بضجيج إضافي

ولما كنا نعرف توزيعات الدخل وتغيرات الخرج، فإن معدل تمرير المعلومات من الدخل إلى الخرج يمكن أن يوجد مباشرة من المعلومات المتبادلة بين الخرج والدخل:

$$I(y;x) = h(y) - h(y|x) \qquad (13.13)$$

لاحظ أن الحد الثانـــي في العلاقة (13.13)، هو الأنتروبـــي الشرطي لـــ y مع x معطى، وهو موزع تماماً كـــ $v$ باعتبار $x_i$ معطى (ثوابت). لذا من تعريف الأنتروبـــي

التفاضلي نجد مباشرة:

$$h(y)=\frac{1}{2}\left[1+\log(2\pi\sigma_v^2)\right] \quad (14.13)$$

وبالمثل، يعطى الأنتروبـــي التفاضلي الشرطي لـــ y مع **x** معطى بالعلاقة التالية:

$$h(y|\mathbf{x})=h(v)=\frac{1}{2}\left[1+\log\left(2\pi\sigma_v^2\right)\right] \quad (15.13)$$

بضم المعادلات (14.13) و(15.13) نجد أن المعلومات المتبادلة تعطى بالعلاقة:

$$I(y;\mathbf{x})=\frac{1}{2}\left[\log\left(2\pi\sigma_y^2\right)-\log\left(2\pi\sigma_v^2\right)\right]$$

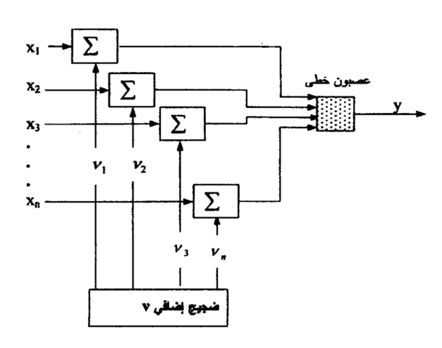
$$=\frac{1}{2}\left[\log\left(\frac{\sigma_y^2}{\sigma_v^2}\right)\right] \quad (16.13)$$

وهكذا، في حالة دخل ضجيج ثابت مع تباين $\sigma_v^2$ نرى أن المعلومات المتبادلة أو معدل حفظ المعلومات يكون أعظمياً عندما يكون $\sigma_y^2$ أعظمياً.

إذن، في حالة قاعدة Hebb المعطاة بالمعادلة (9.13) وافتراضات التوزيعات الطبيعية، كل طبقة في شبكة Linsker تجعل حفظ المعلومات أعظمياً وفق نظرية المعلومات. أعطيت أمثلة بسيطة من قبل Linsker أدت إلى نتائج مشابهة. مثلاً، إذا كان ضجيج معالجة الدخل موزعاً عبـــر وصلات الدخل بحيث أن الدخل إلـــى الوحدة رقـــم i هو $x_i+v_i$ (ومنه $y=\sum_i(x_i+v_i)w_i$)، حيث كل $v_i$ هو توزيع طبيعي مستقل مع تباين عام $\sigma_v^2$ كما هو موضح في الشكل (5.13). نجد أن المعلومات المتبادلة الناتجة تعطى بما يلي:

$$I(y;\mathbf{x})=\frac{1}{2}\left[\log\left(\frac{\sigma_y^2}{\sigma_v^2\sum_i w_i^2}\right)\right] \quad (17.13)$$

في هذه الحالة، أي بمداخل ضجيج ثابتة، تكون المعلومات المتبادلة أعظمية عندما يكون المقدار $\sigma_y^2/\sum w_i^2$ أعظمياً، أي عندما يكون التباين أعظمياً وتكون الأوزان مقيدة على قيم صغيرة. لكن هذا يكافئ أن للوحدة المختارة مجموعة من الأوزان التـــي تنفذ تحليل المركبة الأساسية على نماذج الدخل. وهكذا، بأسلوب مشابه لقاعدة Oja في المعادلة (3.13)، نرى حالة أخرى لتعليم Hebb التـــي تعطي تنظيماً متكيفاً لتحقيق الإنجاز الأمثلي.




**الشكل 5.13**: مداخل ضجيج معالجة متعددد

لتلخيص ما سبق من نتائج نقول إن النموذج المقترح هنا هو شبكة بطبقات مع وصلات تغذية أمامية وتوابع تفعيل خطية. تبنـــي قاعدة التعليم نوع Hebb البسيطة طبقات الوحدات التـــي لها خواص تحليل ملمح أكثر تقدماً بطريقة تزايدية (كلما تقدم تدريب الطبقات). هذه الخواص تشمل الخلايا الخاضعة لشروط مقيدة معينة:

1. جعل تباين تفعيلات خرجها أعظمياً
2. تنفيذ تحليل المركبة الأساسية (استنباط الملمح) على مداخلها
3. حفظ معلومات أعظمية حول تفعيلات الدخل

طبقت الخواص المذكورة آنفاً أيضاً على حالات لم تكن فيها توابع التفعيل خطية بالضرورة.

## 5.13 شبكات التعليم التنافسي بدون معلم

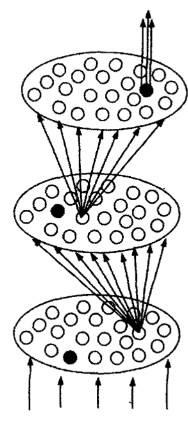
## Unsupervised competitive learning networks

درست شبكات التعليم التنافسي بدون معلم من قبل باحثين كثيرين مثل Grossberg عام 1972[225]، وvon der Malsburg عـــام 1973[226]، وKohonen عـــام 1982[7]،

وZipser و Rumelhart عام 1985[53].

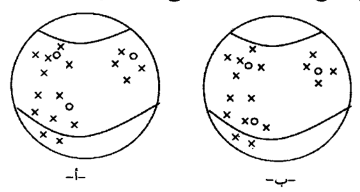
يمكن أن يوصف النموذج التنافسي العام كما يلي: تكون الوحدات منظمة في طبقة أو في طبقات وظيفية عديدة، حيث تكون الوحدات في الطبقة الواحدة مجمعة بتجمعات منفصلة بعضها عن بعض. تحاول كل وحدة ضمن التجمع منع كل الوحدات الأخرى ضمن التجمع لتنافس على مكانة الرابح في منافسة تعتمد مبدأ الرابح يأخذ الكل. الوحدة التـــي تستقبل دخلاً أعظمياً يكون خرجها أعظمياً، وتقاد الوحدات الأخرى إلى خرج يساوي الصفر. تستقبل جميع الوحدات في نفس التجمع نفس المداخل. الشبكة التنافسية بثلاث طبقات موضحة في الشكل(6.13).

تتعلم الوحدة فقط إذا ربحت المنافسة في التجمع، ولا يحدث تعليم بين الوحدات الخاسرة. أنجز التعليم من خلال إعادة توزيع الأوزان على وصلات الدخل إلى الوحدة بحيث تبقى الكمية الكلية للوزن لكل وحدة ثابتة (أي $\sum_i w_{ij} = 1$). هذا يعنـــي أن التعليم ينجز بواسطة إزاحة كميات الأوزان من الوصلات غير الفعالة إلى الوصلات الفعالة للوحدة الرابحة.



**الشكل 6.13**: بنية الشبكة التنافسية العامة

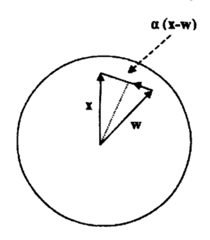
في واحد من تقريبات التعليم العام، تكون قيم أوزان الرابح مزاحة باتجاه شعاع نموذج الدخل. هذه العملية موضحة في الشكل (7.13) حيث افترضنا أن أشعة الدخل والأوزان واحدية معيارية. لذا، تمثل هذه الأشعة بنقاط على كرة بنصف قطر واحدي، تمثل **x** نماذج الدخل وتمثل **o** قيم شعاع الوزن كما هو موضح في الشكل (7.13أ).




**الشكل7.13**: توضيح التعليم التنافسي

في هذا الشكل هناك ثلاث مجموعات من أشعة التدريب موضحة على الكرة مع ثلاثة أشعة وزن قبل أن يحدث أي تدريب. في الشكل (7.13 ب)، بعد إتمام التدريب، تُزاح أشعة الوزن باتجاه مراكز تجمعات أشعة النماذج.

يوضح الشكل (8.13) كيفية تحديث (تعديل القيمة مع تقدم الزمن) شعاع الوزن **w** خلال عملية التعليم. يزاح وزن الوحدة الرابحة باتجاه شعاع نموذج الدخل **x** بإضافة جزء من شعاع الفرق (**x-w**) إلى شعاع الوزن.




**الشكل 8.13**: توضيح تحديث الوزن للتعليم التنافسي

أحد الأمثلة المذهلة للشبكات التنافسية التـــي تتعلم بواسطة التكيف الذاتـــي هو خريطة الملامح ذاتية التنظيم المقترحة من قبل Kohonen بين أعوام 1982[7] و1989[227][228]. وقد استلهم هذا العمل من النتائج المبكرة المنشورة من قبل Grossberg عام 1972[225]، والعمل على التعليم التنافسي الذي ابتكر من قبل von der Malsburg عام 1973[226].

تتكيف شبكات خريطة الملامح ذاتية التنظيم ذاتياً مع نماذج تنبيه الدخل **x** الموصوفة بتوزيع احتمالي ما غير معروف $\rho(\mathbf{x})$. يعطي التكيفُ الناتج شبكةً تطبق نماذج الدخل إلى نماذج الخرج مع التماسك الطبولوجي (مع المحافظة الطبولوجية؛ بنية طبولوجية تفرض فيما بين الوحدات، سيوضَّح معناها جيداً لاحقاً). يستمر التطبيق، مع المحافظة الطبولوجية، الذي سيعكس توزيع احتمال مجتمع الدخل. تبقى خاصية المحافظة على الطبولوجية محققة حتـــى عندما تنفذ خريطة الملامح ذاتية التنظيم تخفيضاً في بعد فراغ الملامح.

استُعملت شبكات خريطة الملامح ذاتية التنظيم في عدد من المسائل كالتكميم الشعاعي، وضغط المعطيات، والأمثلية التركيبية، والتحكم بالربوت، وتعرف الأشكال وتمييزها، وسنقوم بعرض بعض هذه المسائل فيما بعد.

وقبل ذلك، ولفهم أفضل لعملية تكيف خريطة الملامح ذاتية التنظيم، سنصف التكميم الشعاعي. يمكن أن ينظر إلى شبكات التكميم الشعاعي كحالة خاصة من شبكات خريطة الملامح ذاتية التنظيم. ولكن دعنا الآن نعرج على بعض الشبكات التنافسية البسيطة الخاصة ذات الأوزان الثابتة كشبكة الأعظمية (MAXNET)، وشبكة القبعة المكسيكية (Maxican Hat Network)، وشبكة هامنغ (Hamming).

## 6.13 الشبكات التنافسية ذات الأوزان الثابتة
## Fixed-weight competitive networks

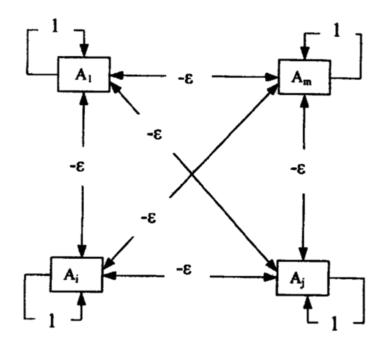
### 1.6.13 شبكة الأعظمية MAXNET

اقترح Lippmann هذه الشبكة عام 1987[5] وهي أنموذج خاص عن الشبكات العصبونية المبنية على التنافس. يمكن أن تستعمل شبكة الأعظمية لانتقاء العقدة التـــي لها الدخل الأكبر.

تتألف هذه الشبكة الجزئية من m عقدة متصلة فيما بينها داخلياً اتصالاً كاملاً، مع أوزان متساوية على جميع الوصلات. ليست هناك خوارزمية تعليم لشبكة الأعظمية؛ لأن أوزانها مثبتة. يوضح الشكل (9.13) بنية هذه الشبكة.

يعطى تابع تفعيل وحدات الشبكة بالعلاقة التالية:

$$f(net) = \begin{cases} net & net > 0 \\ 0 & net \le 0 \end{cases}$$




**الشكل9.13**: شبكة الأعظمية

يمكن وصف عمل الشبكة من خلال الخوارزمية التالية:

1. ضع التفعيلات الأولية والأوزان، افترض $0 < \varepsilon < 1/m$ (m عدد عقد الشبكة) دخل العقدة $A_j$ يساوي $a_j(0)$

$$w_{ij} = \begin{cases} 1 & i = j \\ -\varepsilon & i \ne j \end{cases}$$

2. مادام شرط التوقف غير محقق كرر الخطوات من 3 إلى 5
3. حدِّث تفعيل كل وحدة في حالة $j = 1,2...,m$ وفق المعادلة التالية:

$$a_j^{new} = f\left[a_j^{old} - \varepsilon \sum_{k \neq j} a_k^{old}\right] \qquad (18.13)$$

4. خزن التفعيلات لاستعمالها في التكرار التالي:

$$a_j^{old} = a_j^{new} \quad , \quad j = 1,2,\cdots,m$$

5. اختبر شرط التوقف:

**إذا كان** هناك أكثر من عقدة واحدة لها تفعيل غير صفري استمر **وإلا** توقف.

لاحظ، في الخطوة 3، أن دخل التابع f هو ببساطة الدخل الكلي للعقدة Aj من جميع العقد بما في ذلك العقدة نفسها (وصلة التغذية العكسية الذاتية). هناك بعض التحذيرات والاحتياطات يجب مراعاتها خلال عمل الشبكة لمعالجة الحالة عندما يكون لعقدتين أو أكثر نفس الدخل الأعظمي.

لتوضيح عمل الشبكة سنناقش هذا المثال البسيط، ليكن لدينا الشبكة السابقة المؤلفة من أربع عقد بأوزان مخمدة $\varepsilon = 0.2$، حيث ستكون قيم التفعيلات الأولية (إشارات الدخل) كما يلي:

$$a_1(0) = 0.2\,, a_2(0) = 0.4\,, a_3(0) = 0.6\,, a_4(0) = 0.8$$

تفعيلات الشبكة المكتشفة عند كل تكرار هي التالي:

$a_1(1) = 0.0 \quad a_2(1) = 0.08 \quad a_3(1) = 0.320 \quad a_4(1) = 0.560$

$a_1(2) = 0.0 \quad a_2(2) = 0.00 \quad a_3(2) = 0.192 \quad a_4(2) = 0.480$

$a_1(3) = 0.0 \quad a_2(3) = 0.00 \quad a_3(3) = 0.096 \quad a_4(4) = 0.442$

$a_1(4) = 0.0 \quad a_2(4) = 0.00 \quad a_3(4) = 0.008 \quad a_4(4) = 0.422$

$a_1(5) = 0.0 \quad a_2(5) = 0.00 \quad a_3(5) = 0.000 \quad a_4(5) = 0.421$

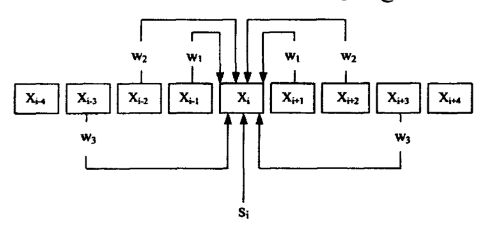
ليس من الضروري، عموماً، تخزين كل القيم السابقة، إذ تلزم عادة تفعيلات الخطوة السابقة فقط كما هو موضح في الخوارزمية.

### 2.6.13 شبكة القبعة المكسيكية Maxican Hat network

اقترح Kohonen هذه الشبكة عام 1989[227]، وهي شبكة جزئية معززة مغايرة لشبكة الأعظمية السابقة. كل عصبون فيها متصل بوصلات مهيجة (أوزان موجبة) إلى عدد من الوحدات "المتجاورة المتعاونة"، التـــي يمكن أن تكون متقاربة جداً.

وكل عصبون يتصل أيضاً بوصلات مخمدة (أوزان سالبة) إلى عدد من الوحدات المتجاورة المتنافسة التـــي تكون بعيدة بعضها عن بعض نوعاً ما. يمكن أن يكون هناك أيضاً عدد من العصبونات التـــي ما تزال أكثر بعداً بحيث تكون منفصلة (أوزان تساوي الصفر). جميع هذه الوصلات تكون ضمن طبقة خاصة من الشبكة العصبونية، وهكذا كما في حالة شبكة الأعظمية، تَستقبل العصبونات إشارة خارجية بالإضافة إلى إشارات الوصلات الداخلية هذه.

يكرر نموذج الوصلات الداخلية تماماً لكل عصبون فـــي الطبقة، كما هو موضح في الشكل(10.13) لنموذج وصل الوحدة $X_i$.




**الشكل 10.13**: الوصلات الداخلية للوحدة $X_i$ في شبكة القبعة المكسيكية

لتسهيل الوصف، صُفّت العصبونات وكأنها مرتبة خطياً بوصلات موجبة بين الوحدة $X_i$ والوحدات المجاورة بمكان أو مكانين على كلا الطرفين (الجيران المتعاونين القريبين مكانياً)، والوصلات السالبة تكون مشاهدة للوحدات في المكان الثالث على كلا الطرفين (الجيران المتباعدين والمتنافسين). يمكن أن يتغير حجم منطقة التعاون (الوصلات الموجبة) وحجم منطقة التنافس (الوصلات السالبة)، ويمكن أن تكون أحجام هذه المناطق مقادير متناسبة مع الأوزان الموجبة والسالبة ومع طبولوجية المناطق (خطية، مستطيلة، سداسية،.. الخ). ينجز تعزيز التغاير للإشارات $S_i$ المستقبلة بالوحدة $X_i$ بواسطة التكرار لخطوات زمنية عديدة.

يعطى تفعيل الوحدة $X_i$ عند الزمن t بالعلاقة:

$$x_i(t) = f\left[ s_i(t) + \sum_k w_k x_{i+k}(t-1) \right] \qquad (19.13)$$

حيث الحدود في المجموع هي إشارات مثقلة من الوحدات الأخرى (جوار تعاونـــي أو تنافسي) عند خطوة زمنية سابقة.

في المثال الموضح في الشكل (10.13)، سيكون الوزن $v_k$ من الوحدة $X_i$ إلى الوحدة $X_{i+k}$ موجباً في حالة $k = -2,-1,0,1,2$، وسالباً في حالة $k = -3,3$، وصفراً في غير ذلك من الوحدات. تحقق الوصلات الداخلية لشبكة القبعة المكسيكية منطقتين متناظرتين حول كل وزن مفرد. ستكون أوزان الوصلات ضمن المنطقة التي هي أقرب (الأوزان بين الوحدة النموذجية $X_i$ والوحدات $X_{i+1}$ و$X_{i+2}$ و$X_{i-1}$ و$X_{i-2}$ مثلاً) موجبة (وغالباً تعطى قيمة متساوية). تظهر هذه الأوزان كـــ $w_1$ و$w_2$ في الشكل (10.13). وستكون الأوزان بين الوحدة $X_i$ والوحدات $X_{i+3}$ و$X_{i-3}$ سالبة (تظهر كـــ $w_3$ في الشكل)، والوحدة $X_i$ ليست موصلة مع الوحدات $X_{i+4}$ و$X_{i-4}$ في هذه البنية.

في هذا الشكل التوضيحي، ستكون الوحدات ضمن نصف قطر 2 (يحدد مساحة كل منطقة جوار) إلى الوحدة النموذجية $X_i$ موصلةً بأوزان موجبة، وتكون الوحدات ضمن نصف القطر 3، ولكن خارج نصف القطر للوصلات الموجبة، موصلة بأوزان سالبة، وستكون الوحدات التي هي أبعد من ثلاثة وحدات غير موصلة.

### 1.2.6.13 خوارزمية تعليم شبكة القبعة المكسيكية

الخوارزمية المعطاة هنا مشابهة لتلك التـــي قدمها Kohonen عام 1989[227]، وقبل استعراض خطوات هذه الخوارزمية سنعرف بعض المصطلحات:

$R_1$ : نصف قطـــر منطقة الوصلات الداخلية الكلية، حيث $X_i$ موصلة إلى الوحدات $X_{i+k}$ و $X_{i-k}$ في حالة $k = 1,2,\cdots,R_1$.

$R_2$ : نصف قطر منطقة التعزيز الموجب فقط، حيث $R_2 < R_1$

$w_k$ : الوزن على الوصلات الداخلية بين $X_i$ والوحدات $X_{i+k}$ و $X_{i-k}$ ويكون $w_k > 0$ في حالة $0 \le k \le R_2$ و$w_k < 0$ في حالة $R_2 < k \le R_1$

**X**: شعاع التفعيل

**x_old**: شعاع التفعيلات عند الخطوة الزمنية السابقة

**t_max**: العدد الكلي لتكرارات التعزيز المغاير

s: إشارة خارجية

كما قدمنا، الخوارزمية موافقة لإشارة خارجية معطاة فقط في التكرار الأول (الخطوة 1) لتكرارات التعزيز المغاير.

1. تعطـــى الوسطاء قيماً أولية كما هـــو مـــرغوب: t_max، R1، R2، وستكـــون الأوزان الأوليـــة: $w_k = C_1$ في حالة $k = 0,1,2,\cdots,R_2$ $(C_1 > 0)$، $w_k = C_2$ في حالة $k = R_2 + 1,\cdots,R_1$ $(C_2 < 0)$ ووضع x_old بقيمة أولية صفرية

2. تقديم الإشارة الخارجية s إلى دخل الشبكة:

$$\mathbf{x} = \mathbf{s}$$

تخزين التفعيلات في مصفوفة x_old (في حالة $i = 1,2,...,n$):

$$x_old_i = x_i$$

ضع عداد التكرار t = 1

3. مادام t أقل من t_max، كرر الخطوات من 4 إلى 8

4. احسب دخل الشبكة ($i = 1,2,...,n$):

$$x_i = C_1 \sum_{k=-R_2}^{R_2} x_old_{i+k} + C_2 \sum_{k=-R_1}^{-R_2-1} x_old_{i+k} + C_2 \sum_{k=R_2+1}^{R_1} x_old_{i+k} \quad (20.3)$$

5. طبق تابع التفعيل (تابع خطي بين الصفرو x_max بميل يساوي الواحد):

$$x_i = \min(x_\max, \max(0, x_i)) \quad , \quad i = 1,2,\cdots,n$$

6. خزن التفعيلات الحالية في x_old:

$$x_old_i = x_i$$

$$i = 1,2,...,n$$

7. زيادة عداد التكرار t = t + 1

8. اختبر شرط التوقف:

إذا كان t <t_max استمر وإلا توقف.

سنوضح خوارزمية شبكة القبعة المكسيكية في حالة شبكة بسيطة بسبع وحدات. يعطى تابع التفعيل لوحدات هذه الشبكة بـــ:

$$f(net) = \begin{cases} 0 & net < 0 \\ net & 0 \le net \le 2 \\ 2 & 2 < net \end{cases}$$

1. الوسطاء الأولية:

$$R_1 = 2, \quad R_2 = 1, \quad C_1 = 0.6, \quad C_2 = -0.4$$

2. $t = 0$

الإشارة الخارجية هي:

$$(0.0 , 0.5 , 0.8 , 1.0 , 0.8 , 0.5 , 0.0)$$

ومنه:

$$\mathbf{x} = (0.0 , 0.5 , 0.8 , 1.0 , 0.8 , 0.5 , 0.0)$$

التخزين في **x_old**:

$$\mathbf{x_old} = (0.0 , 0.5 , 0.8 , 1.0 , 0.8 , 0.5 , 0.0)$$

3. $t = 1$، صيغ التحديث المستعملة في الخطوة 4 ستكون كما يلي:

$$x_1 = 0.6x - \text{old}_1 + 0.6x - \text{old}_2 - 0.4x - \text{old}_3$$
$$x_2 = 0.6x - \text{old}_1 + 0.6x - \text{old}_2 + 0.6x - \text{old}_3 - 0.4x - \text{old}_4$$
$$x_3 = -0.4x - \text{old}_1 + 0.6x - \text{old}_2 + 0.6x - \text{old}_3 + 0.6x - \text{old}_4 - 0.4x - \text{old}_5$$
$$x_4 = -0.4x - \text{old}_2 + 0.6x - \text{old}_3 + 0.6x - \text{old}_4 + 0.6x - \text{old}_5 - 0.4x - \text{old}_6$$
$$x_5 = -0.4x - \text{old}_3 + 0.6x - \text{old}_4 + 0.6x - \text{old}_5 + 0.6x - \text{old}_6 - 0.4x - \text{old}_7$$
$$x_6 = -0.4x - \text{old}_4 + 0.6x - \text{old}_5 + 0.6x - \text{old}_6 + 0.6x - \text{old}_7$$
$$x_7 = -0.4x - \text{old}_5 + 0.6x - \text{old}_6 + 0.6x - \text{old}_7$$

4. $t = 1$

$$x_1 = 0.6(0.0) + 0.6(0.5) - 0.4(0.8) = -0.02$$
$$x_2 = 0.6(0.0) + 0.6(0.5) + 0.6(0.8) - 0.4(1.0) = 0.38$$
$$x_3 = -0.4(0.0) + 0.6(0.5) + 0.6(0.8) + 0.6(1.0) - 0.4(0.8) = 1.06$$
$$x_4 = -0.4(0.5) + 0.6(0.8) + 0.6(1.0) + 0.6(0.8) - 0.4(0.5) = 1.16$$
$$x_5 = -0.4(0.8) + 0.6(1.0) + 0.6(0.8) + 0.6(0.5) - 0.4(0.0) = 1.06$$
$$x_6 = -0.4(1.0) + 0.6(0.8) + 0.6(0.5) + 0.6(0.0) = 0.38$$
$$x_7 = -0.4(0.8) + 0.6(0.5) + 0.6(0.0) = -0.02$$

.5

$$\mathbf{x} = (0.0\ ,\ 0.38\ ,\ 1.06\ ,\ 1.16\ ,\ 1.06\ ,\ 0.38\ ,\ 0.0)$$

6 حتـــى 8 تنفذ الحسابات في التكرار التالي

2 = t .4

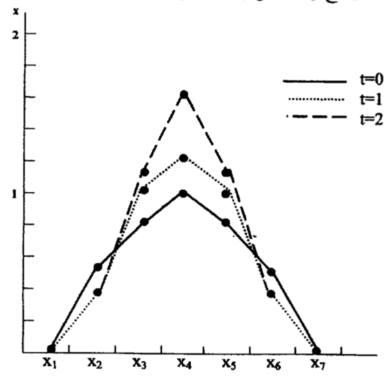
$$x_1 = 0.6(0.0) + 0.6(0.38) - 0.4(1.06) = -0.196$$
$$x_2 = 0.6(0.0) + 0.6(0.38) + 0.6(1.06) - 0.4(1.16) = 0.39$$
$$x_3 = -0.4(0.0) + 0.6(0.38) + 0.6(1.06) + 0.6(1.16) - 0.4(1.06) = 1.14$$
$$x_4 = -0.4(0.38) + 0.6(1.06) + 0.6(1.16) + 0.6(1.06) - 0.4(0.38) = 1.66$$
$$x_5 = -0.4(1.06) + 0.6(1.16) + 0.6(1.06) + 0.6(0.38) - 0.4(0.0) = 1.14$$
$$x_6 = -0.4(1.16) + 0.6(1.06) + 0.6(0.38) + 0.6(0.0) = 0.39$$
$$x_7 = -0.4(1.06) + 0.6(0.38) + 0.6(0.0) = -0.196$$

.5

$$\mathbf{x} = (0.0\ ,\ 0.39\ ,\ 1.14\ ,\ 1.66\ ,\ 1.14\ ,\ 0.39\ ,\ 0.0)$$

6 حتـــى تنفذ الحسابات في التكرار التالي.

نموذج التفعيلات موضح في الشكل (11.13) في حالة t = 0,1,2.




**الشكل 11.13:** نتائج مثال القبعة المكسيكية

### 3.6.13 شبكة هامنغ

شبكة هامنغ (Lippmann عام 1987[5] وDARPA عام 1988[183]) هي شبكة مصنف الأرجحية العظمى التـــي يمكن أن تستعمل لتحديد أي من الأشعة الأنموذج المتعددة يكون مشابهاً أكثر لشعاع الدخل ببعد n. تعين الأشعة الأنموذج أوزان الشبكة.

يعطى قياس التشابه بين شعاع الدخل وأشعة الأنموذج المخزنة بـــ n ناقصاً مسافة هامنغ بين الأشعة. تذكر أن مسافة هامنغ بين شعاعين هي عدد المركبات المختلفة في كلا الشعاعين. ففي أشعة ثنائية القطبية x وy:

$$\mathbf{x}.\ \mathbf{y} = a - d \qquad (21.13)$$

حيث a عدد المركبات المتشابهة في كلا الشعاعينو d عدد المركبات المختلفة في كلا الشعاعين، وهي مسافة هامنغ. على أية حال، إذا كان n عدد المركبات في الأشعة، فإن:

$$d = n - a$$

و

$$\mathbf{x}.\ \mathbf{y} = 2a - n$$

$$2a = \mathbf{x}.\ \mathbf{y} + n \qquad (22.13)$$

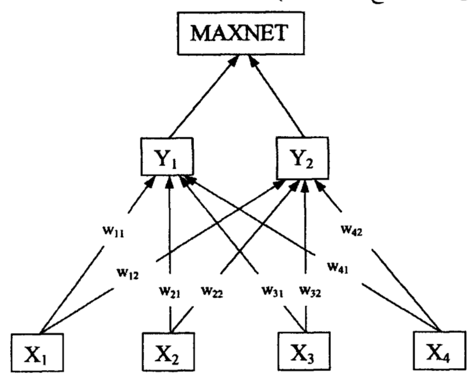
فإذا اخترنا الأوزان بحيث تقع في النصف الأول من الشعاع الأنموذج، وبوضع قيمة الانحياز مساوية n/2، ستجد الشبكة الوحدة ذات الأنموذج الأقرب ببساطة بإيجاد الوحدة ذات دخل الشبكة الأكبر.

تستعمل شبكة هامنغ شبكة الأعظمية كشبكة جزئية ضمن بنيتها العلوية (شبكة الأعظمية تعتبر رأس شبكة هامنغ) لإيجاد الوحدة ذات دخل الشبكة الأكبر. تتألف الشبكة السفلية من n عقدة دخل، كل منها متصل مع m عقدة خرج (حيثm عدد الأشعة الأنموذج المخزنة في الشبكة).

تغذى عقد خرج الشبكة السفلية إلى الشبكة العلوية (شبكة الأعظمية) التـــي تحسب الأنموذج الأقرب الذي يلائم شعاع الدخل. الدخل وأشعة الأنموذج هي ثنائية القطبية.

بنية هذه الشبكة موضحة في الشكل (12.13)، وذلك بافتراض أن أشعة دخل ببعد 4

موزعة إلى فئات بحيث تنتمي إلى أحد صفين اثنين. إذا كان لدينا m شعاع أنموذج ثنائية القطبية e(1),e(2),....,e(m)، عندها تستطيع شبكة هامنغ إيجاد الأنموذج الذي سيكون الأقرب إلى شعاع الدخل ثنائي القطبية x. يعطي دخل الشبكة $y(net_j)$ للوحدة $y_j$ (دخل الوحدة $y_j$) عدد المركبات المتشابهة في كلا شعاع الدخل وشعاع الأنموذج e(j) في الوحدة $Y_j$(n ناقص مسافة هامنغ بين الشعاعين).




**الشكل 12.13**: شبكة هامنغ

قبل مناقشة خوارزمية الشبكة سنعتمد المصطلحات التالية:

n: عدد عقد الدخل، أي عدد المركبات لأي شعاع دخل.

m: عدد عقد الخرج، أي عدد الأشعة الأنموذج.

e(j): شعاع الأنموذج رقم j:

$$e(j) = (e_1(j), e_2(j), \cdots, e_i(j), \cdots, e_n(j))$$

وستكون خوارزمية عمل الشبكة على النحو التالي:

1. تخزين m شعاع أنموذج، ووضع القيم البدائية للأوزان:

$$w_{ij} = \frac{1}{2} e_i(j), (i = 1,2,\cdots,n\,; j = 1,2,\cdots,m)$$

ووضع القيم البدائية للانحيازات:

$$b_j = \frac{1}{2} n, (j = 1,2,\cdots,m)$$

2. لكل شعاع **x** نفذ الخطوات من 3 إلى 5:

3. احسب دخل الشبكة لكل وحدة Yj:

$$y(net_j) = b_j + \sum_i x_i w_{ij} \quad ,(j = 1,2,\cdots,m)$$

4. التفعيلات الأولية لشبكة الأعظمية:

$$y_j(0) = y(net_j) \quad ,(j = 1,2,\cdots,m)$$

5. كرر شبكة الأعظمية الحساب حتـــى تجد الأنموذج الأكثر ملاءمة.

سنحاول الآن تنفيذ المثال التالي لكي نفهم أفضل خطوات هذه الخوارزمية. ليكن لدينا الشعاعين الأنموذجين التاليين:

$$\mathbf{e}(1) = (1, -1, -1, -1)$$

$$\mathbf{e}(2) = (-1, -1, -1, 1)$$

يمكن استعمال شبكة هامنغ لإيجاد الأنموذج الأقرب إلى نماذج الدخل ثنائية القطبية التالية:

$$(1, -1, -1, -1) \quad (1, 1, -1, -1)$$

$$(-1, -1, 1, 1) \quad (-1, -1, -1, 1)$$

1. خزن m شعاع أنموذج في الأوزان:

$$\mathbf{W} = \begin{bmatrix} +0.5 & -0.5 \\ -0.5 & -0.5 \\ -0.5 & -0.5 \\ -0.5 & +0.5 \end{bmatrix}$$

القيم الأولية للانحيازات: $b_1 = b_2 = 2$

2. في حالة شعاع الدخل الأول $\mathbf{x} = (1, 1, -1, -1)$، كرر الخطوات من 3 إلى 5

.3

$$y(net_1) = b_1 + \sum_i x_i w_{i1} = 2 + 1 = 3$$
$$y(net_2) = b_2 + \sum_i x_i w_{i2} = 2 - 1 = 1$$

تمثـــل هـــذه القيـــم قيـــاس تشابه هامنغ بسبب أن $\mathbf{x} = (1, 1, -1, -1)$ يتفـــق مـــع $e(1) = (1, -1, -1, -1)$ في المركبات الأولى والثالثة والـــرابعة ولكنـــه يتفـــق مع $e(2) = (-1, -1, -1, 1)$ في المركبة الثالثة فقط.

.4

$$y_2(0) = 1 \text{ و } y_1(0) = 3$$

5. باعتبار $y_1(0) > y_2(0)$ فإن شبكة الأعظمية ستجد أن الوحدة $Y_1$ لها الأنموذج الأكثر ملائمة لشعاع الدخل $\mathbf{x} = (1,1,-1,-1)$.

1. في حالة شعاع الدخل الثانـــي $\mathbf{x} = (1,-1,-1,-1)$، كرر الخطوات من 2 إلى 4

.2

$$y(net_1) = b_1 + \sum_i x_i w_{i1} = 2 + 2 = 4$$
$$y(net_2) = b_2 + \sum_i x_i w_{i2} = 2 - 0 = 2$$

تمثـــل هذه القيـــم قيـــاس تشابـــه هامنـــغ بسبب أن $\mathbf{x} = (1, -1, -1, -1)$ يتفق مع $e(1) = (1, -1, -1, -1)$ في المركبات الأربعة، ولكنه يتفق مع $e(2) = (-1, -1, -1, 1)$ في المركبة الثانية الثالثة فقط.

.3

$$y_2(0) = 2 \text{ و } y_1(0) = 4$$

4. باعتبار $y_1(0) > y_2(0)$ فإن شبكة الأعظمية ستجد أن الوحدة $Y_1$ لها الأنموذج الأكثر ملائمة لشعاع الدخل $\mathbf{x} = (1, -1, -1, -1)$.

1. في حالة شعاع الدخل الثالث $\mathbf{x} = (-1, -1, -1, 1)$، كرر الخطوات من 2 إلى 4

.2

$$y(net_1) = b_1 + \sum_i x_i w_{i1} = 2 + 0 = 2$$

$$y(net_2) = b_2 + \sum_i x_i w_{i2} = 2 + 2 = 4$$

تمثـــل هـــذه القيـــم قياس تشابه هامنغ بسبب أن $\mathbf{x} = (-1, -1, -1, 1)$ يتفـــق مـــع $e(1) = (1, -1, -1, -1)$ في المركبات الثانية والثالثة، ولكنه يتفق مع $e(2) = (-1, -1, -1, 1)$ في المركبة الرابعة فقط.

.3

$$y_1(0) = 2 \text{ و } y_2(0) = 4$$

4. باعتبار $y_2(0) > y_1(0)$ فإن شبكة الأعظمية ستجد أن الوحدة $Y_2$ لها الأنموذج الأكثر ملاءمة لشعاع الدخل $\mathbf{x} = (-1, -1, -1, 1)$.

1. في حالة شعاع الدخل الرابع $\mathbf{x} = (-1, -1, 1, 1)$، كرر الخطوات من 2 إلى 4

.2

$$y(net_1) = b_1 + \sum_i x_i w_{i1} = 2 - 1 = 1$$
$$y(net_2) = b_2 + \sum_i x_i w_{i2} = 2 + 1 = 3$$

تمثـــل هذه القيـــم قيـــاس تشابه هامنغ بسبب أن $\mathbf{x} = (-1, -1, 1, 1)$ يتفـــق مـــع $e(1) = (1, -1, -1, -1)$ في المركبات الثانية فقط، ولكنه يتفق مع $e(2) = (-1, -1, -1, 1)$ في المركبة الأولى والثانية والرابعة.

.3

$$y_1(0) = 1 \text{ و } y_2(0) = 3$$

4. باعتبار $y_2(0) > y_1(0)$ فإن شبكة الأعظمية ستجد أن الوحدة $Y_2$ لها الأنموذج الأكثر ملاءمة لشعاع الدخل $\mathbf{x} = (-1, -1, 1, 1)$.

## 7.13 التكميم الشعاعي Vector Quantization(VQ)

نحتاج في تطبيقات عديدة، مثل تمييز إشارة الكلام أو تعرف الأشكال ومعالجة الصور، إلى تخزين كميات ضخمة من المعطيات وبثها عبر أقنية الاتصالات. مثلاً، يتطلب التحليل الدقيق لصورة واحدة أكثر من 1000× 1000 بايت (byte) من المعطيات، كل منها توافق قيمة شدة مستوى رمادي عنصر صورة (pixel) واحد. عندما يكون هناك عدة صور تحت المعالجة، فإن كمية المعطيات المعالجة يمكن أن تصبح ضخمة جداً. نموذجياً، هناك زيادة مفرطة في المعطيات في هذا النوع من المعالجات. وهناك أجزاء ضخمة من الصورة مثل الخلفية السماوية أو أشياء متجانسة أخرى سيكون لها نفس مستويات الشدة تقريباً أو نماذج التركيب المتكررة.

عندما تكون المعطيات المتقاربة بنفس القيم تقريباً، فإن نوعاً ما من ضغط المعطيات أو الترميز يمكن أن ينجز لتقليل الكمية الكلية للمعطيات المعالجة. مثلاً، يمكن أن تجمع عناصر الصورة المتجاورة (أحرف أو أرقام) بنفس القيم تقريباً وتخصص بدليل مفرد أو رمز واحد.

من جهة أخرى، فإن النموذج الذي يمكن أن يجمع في واحد من عدد محدود من الصفوف يمكن أن يخصص بدليل شعاع أولي لذاك الصف. إن طول الترميز الناتج عن عملية ضغط المعطيات يمكن أن يكون أصغر بكثير من الكمية الكبيرة الأصلية للمعطيات. والإنقاص المحقق في عرض حزمة البث وفي معالجة المعلومات وكمية تخزينها يمكن أن يبلغ أقصى مدى له من30-80% .

تسمح بعض طرائق الضغط باستعادة كاملة للمعطيات الأصلية وبعضها الآخر لا يسمح بذلك. في الحالة الأخيرة، يكون هناك تسوية أو موازنة بين كمية الضياع في التحليل نتيجة عملية الضغط والإنقاص في كمية المعطيات؛ بعبارة أخرى، يؤخذ بالحسبان ما نربحه من عملية الضغط وما نخسره نتيجة لهذا الضغط.

إذا كان العدد الكلي لكلمات الرمز k صغيراً، فإن صفوفاً أو مستويات شدة أقل يمكن أن تمثل، وقد تصبح عملية الاستعادة الكاملة مستحيلة. إذا كان k كبيراً، فإنه سيكون هناك ضياع قليل في التحليل، لكن سيتحقق فقط تخفيض (ضغط) صغير في المعطيات. من الواضح أن الاختيار الأفضل لـــ k يعتمد على المسألة المعالجة.

لقد طورت تقنيات عديدة لضغط المعطيات (Devijver وKittler عام 1982[28])، لكن أفضلها هي التـــي اعتمدت على معرفة ما بتوزيع الاحتمال $\rho(\mathbf{x})$ الذي تستمد منه نماذج الدخل $\mathbf{x}$.

تستفيد خطط الترميز الفعالة من حسنات التكرارات النسبية لحدوث نماذج الدخل بواسطة تخصيص كلمات رمز أقصر للنماذج التـــي تحدث بتكرار كثير. أحد أكثر الأمثلة الشائعة للترميز هو رمز Morse، حيث استعملت شرطة واحدة (dash) لترميز الحرف الذي يتواتر كثيراً في الأحرف الإنكليزية، وهو الحرف E.

وباستعمال نظرية المعلومات، من الممكن دائماً ابتكار خطة ترميز فعالة كثيراً عندما تكون إحصائيات المنبع معروفة. إما إذا كانت المعرفة المتوفرة عن توزيع المنبع قليلة، وخاصة، عندما يكون التوزيع غير خطي أبداً، فيمكن أن تكون هناك طرق أخرى فعالة أكثر.

الطريقة الوحيدة للضغط التـــي طبقت بنجاح على بنـــى الشبكات العصبونية الصنعية هي التكميم الشعاعي. يمكن اللجوء إلى تقنية ضغط المعطيات بالشبكات العصبونية الصنعية عندما تتوفر لدينا معرفة قليلة عن توزيع المنبع. درس هذا التقريب وأثبت مقدرة جيدة مقارنة مع تقنيات ضغط المعطيات الأخرى من قبل Kohonen عام 1988[179].

تكميم الشعاع هو عملية تطبيق الأشعة $\mathbf{x}$، التـــي تكون عادة أشعة مستمرة بقيم حقيقية، من جملة مُولدة A، حيث $A \in \mathbf{R}^n$، على الشعاع المرجع (شعاع وزن وحدة الخرج) الأقرب $\mathbf{w}_i$ المنتمي إلى الجملة المولدة B، حيث $B \in \mathbf{R}^m$. بكلمات أخرى، ستحول أشعة الدخل $\mathbf{x}$ ذات البعد n إلى واحد من عدد محدود من الصفوف، حيث يمثل كل صف بواسطة كلمة رمز أو شعاع أولي $\mathbf{w}_i$، $I=1,2,\ldots,m$. الدليل $i$ في البعد $m < n$ يصبح مؤشر صف لـــ $\mathbf{x}$.

تطبيق التكميم الشعاعي $f: A \to B$ هو تطبيق الجوار الأقرب، حيث يمكن أن يعرف الأقرب بطرق مختلفة. نموذجياً، هو المسافة الإقليدية أو تابع كلفة مثل تشويه مربع الخطأ المعرف بما يلي:

$$d(\mathbf{x},\mathbf{w}_i) = \|\mathbf{x}-\mathbf{w}_i\|^2 = \sum_{j=1}^{n}\left(x_j - w_{ij}\right)^2 \qquad (23.13)$$

حيث $\mathbf{w}_i$ الأقرب إلى $\mathbf{x}$ هو ذو التشويه الأصغري $\mathbf{d}$.

تعتبر شبكات التكميم الشعاعي هنا شبكات تنافسية حيث m وحدة خرج تتنافس لتمثيل نماذج الدخل كما وصف في المقاطع السابقة. وتصبح الوحدة ذات شعاع الوزن الأقرب إلى $\mathbf{x}$ رابحة في منافسة الرابح يأخذ الكل.

تقوي عصبونات التنافس تهيجها الخاص من خلال وصلة التغذية العكسية الذاتية وتمنع وحدات المنافسة الأخرى من خلال الوصلات الجانبية. وتربح المنافسة الوحدةُ ذات تهييج الدخل الأقوى. شبكة التكميم الشعاعي موضحة في الشكل (13.13). النموذج الكامل يأخذ بالحسبان مداخل خارجية $\mathbf{x}^T\mathbf{w}_j$ إلى الوحدة رقم j بالإضافة إلى تغذية عكسية داخلية فيما بين الوحدات.

يمكن أن توصف ديناميكية النظام بمجموعة من المعادلات التفاضلية كتابع لقيم تفعيل $y_j$ للوحدة رقم $j$ $(j = 1,2, ...,m)$ المعطاة كما يلي:

$$\frac{dy_j}{dt} = \sum_i w_{ij} x_i + \sum_{k \in S_j} v_{kj} y_k - h(y_j) \qquad (24.13)$$

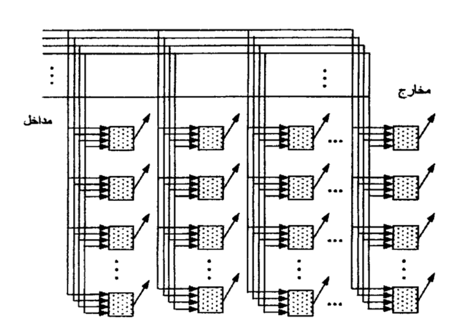
حيث $x_i$ مداخل، $i = 1,2,...,n$، و $w_{ij}$ الوزن على الوصلة بين الدخل رقم i والوحدة رقم j، و $v_{kj}$ الوزن على الوصلة الداخلية من خرج الوحدة رقم k إلى دخل الوحدة رقم j، و$S_j$ مجموعة الوحدات التـــي لها وصلات مع الوحدة j، و $h(y_j)$ حد التسرب غير الخطي الذي يؤخذ بالحسبان في مجموع المؤثرات مثل الإشباع والتسرب والتفريغ. فقط الأوزان $w_{ij}$ على وصلات الدخل تكون قابلة للتكييف، على حين أن الأوزان $v_{kj}$ على وصلات التغذية العكسية تكون مثبتة.

المعادلات التفاضلية التـــي تصف عملية التعليم هي:

$$\frac{dw_{ij}}{dt} = \alpha(x_i - w_{ij}), \quad y_j = 1 \qquad \text{(أ25.13) ( الرابح)}$$

$$\frac{dw_{ij}}{dt} = 0 \qquad y_j \neq 1 \qquad \text{(ب25.13)}$$

حيث $\alpha$ معدل التعليم.




**الشكل 13.13:** شبكة التكميم الشعاعي

عندما يكون مجموع قيم الأوزان ثابتاً لكل وحدة، ليكن $\sum w_{ij} = 1$ ، ونفس القيمة لكل الوحدات، وعندما تكون قيم الدخل معيارية، $\|\mathbf{x}\| = 1$ ، عندها يمكن استعمال تقريب مختصر لتحديد الوحدة الرابحة (Kohonen عام 1984[229]). يستعمل هذا التقريب النظيم الإقليدي كقياس للقرب بين أشعة الدخل وأوزان الوحدات. وهكذا، فإن الوحدة c مع شعاع الوزن $w_c$ تربح المنافسة عندما:

$$\|\mathbf{w}_c - \mathbf{x}\| = \min_i \|\mathbf{w}_i - \mathbf{x}\| \qquad (26.13)$$

عندما تكون أطوال شعاع الوزن مثبتة والمداخل معممة، فإن الجداء السلمي الأعظمي $\mathbf{x}^T\mathbf{w}_j$ يكون مكافئاً للمعادلة (26.13).

ينفذ التعليم في هذه الشبكات فقط بواسطة الوحدة الرابحة وبطريقة ما بحيث يكون شعاع الوزن $\mathbf{w}_c$ للوحدة الرابحة مزاحاً باتجاه نموذج الدخل $\mathbf{x}$ (الشكل(12.13)). تعطى قاعدة التحديث المتكيف بالعلاقات التالية:

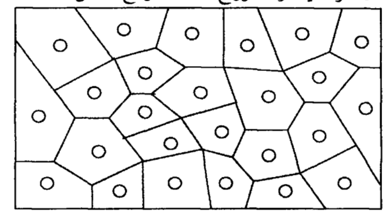
من أجل الوحدة الرابحة c (27.13)

$$\mathbf{W}_c^{new} = \mathrm{W}_c^{old} + a\,(\mathbf{X} - \mathbf{W}_c)$$
$$\mathbf{W}_i^{new} = \mathrm{W}_i^{old}$$

ومن أجل $i \neq c$

حيث $\alpha > 0$ معدل التعليم الذي يمكن أن يكون ثابتاً أو متناقصاً كلما تقدم التعليم. العملية المعرفة بالمعادلات (26.13) و(27.13) مكافئة لـ k متوسط تجمع وبأسلوب متقارب، تجزأ الأوزان $w_i$ فراغ النموذج إلى مناطق موصوفة بواسطة الترصيع بالمضلعات لـ Voronoi. يوضح الشكل (14.13) التجزيء الأولي ثنائي البعد والأشعة الأولية المرجعية الموافقة.

في فراغ ببعد n، يعطى التجزيء غير الخطي بواسطة أسطح محددة (بحدود فصل) لها كثافات خلوية داخلية، والذي يقرب توزيع احتمال نموذج الدخل.



**الشكل 14.13**: تجزيء فراغ النموذج بترصيع مضلع ثنائي البعد

أي نموذج يسقط ضمن منطقة معطاة سيخصص بدليل يعرف خلية الترصيع الخاصة. توافق مجموعات من خلية واحدة أو أكثر تعيين صفوف مختلفة. المناطق المتقطعة في $\mathbf{R}^n$ المعرفة بواسطة (26.13) و(27.13) تصل إلى تكميم أمثلي تقريباً للفراغ الشعاعي. وهذا يعنـــي، أن عدد النماذج المصنفة خطأ سيكون أصغرياً، لأن مسافة أي نقطة ضمن الخلية ستكون أقرب إلى النقطة الأولية منها إلى أي نقطة أولية أخرى.

لوحظ أن هذا الشكل من تجزيء التكميم الشعاعي يقرب بدقة الطريقة النظرية المبنية على أسطح قرار (فصل) بايز( Kphonenعام 1988[179]).

لتلخيص ما سبق، يمكن وصف التعليم في شبكة التكميم الشعاعي كما يلي:

1. وضع الأوزان $w_{ij}$ بقيم عشوائية صغيرة (لأول m قيمة نموذج). كرر الخطوات من 2-3 حتـــى تستقر الشبكة .

2. إيجاد الوحدة الأولية لتمثيل **x** بحساب:

$$\|\mathbf{w}_c - \mathbf{x}\| = \min_i \|\mathbf{w}_i - \mathbf{x}\|$$

3. تحديث أشعة الأوزان وفقاً لـ :

من أجل الوحدة الرابحة c

$$\mathbf{W}_c^{new} = \mathbf{W}_c^{old} + a\,(\mathbf{X} - \mathbf{W}_c)$$

$$\mathbf{W}_i^{new} = \mathbf{W}_i^{old}$$

ومن أجل $i \neq c$

النموذج السابق لتكميم الشعاع هو نموذج من التعليم بدون معلم أو التكيف الذاتي. ليس هناك أي هدف منشود أولي معطى لكل دخل.

درست نماذج التعليم بمعلم أيضاً في شبكات التكميم الشعاعي.

## 8.13 النماذج المعدلة للتكميم الشعاعي

**Modified forms of vector quantization**

إن النموذج البسيط للتكميم الشعاعي الموصوف فيما سبق لديه بعض الضعف. فإذا كان التوزيع الأصلي لأشعة الوزن ونماذج الدخل غير منتظم، يمكن لبعض أشعة الوزن ألاّ تختار مطلقاً كرابحة، ومن ثم لن تتعلم مطلقاً.

إذا جمعت بعض أشعة الوزن معاً (لتكوين تجمع) بعيداً عن النماذج، عندئذ يمكن أن يختار شعاع واحد من خلال مراحل التعليم الأولى ويكون مسحوباً باتجاه أشعة الدخل. وستختار المداخل اللاحقة الأخرى نفس الشعاع للتعليم، تاركة الأشعة الأخرى إلى الوراء دون أن تتعلم مطلقاً. لتخفيف تأثير هذه الأنانية والتميُّز، أُدرج نوع من آلية "الضمير" في معادلات التعليم (Di Sieno عام 1988[230]). فالوحدات التـــي تربح تكرارياً يمكن أن تعاقب بإضافة حد انحياز يزيد بأسلوب فعّال مسافة الحساب بالتناسب مع التكرار الذي تربح به الوحدة.

ليكن $p_i$ جزء الزمن الذي تربح خلاله الوحدة $i$ المنافسة. بعدئذ سنعرف $p_i$ بالعلاقة التالية:

$$p_i^{new} = p_i^{old} + b(y_i - p_i^{old}) \qquad (28.13)$$

حيث b ثابت، 0 < b < 1. إذا كانت $z_i$ تمثل الوحدة الرابحة، عندئذ سيقدم حد الانحياز أو العقوبة $B_i$ لتعديل المنافسة:

$$z_i = 1 \text{ فإن } |\mathbf{w}_i - \mathbf{x}|^2 - B_i \le |\mathbf{w}_j - \mathbf{x}|^2 - B_j \quad (\forall i \ne j) \text{ إذا كان}$$

$$z_i = 0 \text{ وما عدا ذلك فإن } \tag{29.13}$$

حد العقوبة $B_i$ يعطى بـــ:

$$B_i = C(1/n - p_i) \tag{30.13}$$

حيث C هو عامل الانحياز، و n عدد وحدات الشبكة. ينشئ C مسافة ضياع للوحدة حتـــى تستطيع الوصول والدخول في الحل. أخيراً تحدث أوزان الوحدة الرابحة للمنافسة وفقاً للعلاقة التالية :

$$\mathbf{W}_i^{new} = \mathbf{W}_i^{old} + \alpha(\mathbf{X} - \mathbf{W}_i^{old})z_i \tag{31.13}$$

الثابت α هو معدل التعليم، وهو جزء من المسافة التـــي تتحركها الوحدة الرابحة باتجاه شعاع الدخل.

تعتبر آلية الضمير الموصوفة آنفاً طريقة فعالة في تطوير متساوي الاحتمال لملامح أوليات وسط الدخل. لقد أثبت تحسين إنجاز الشبكات المختلفة المستعملة نموذجاً من التعليم التنافسي.

فيما يلي سنلخص التغيرات الأخرى على التكميم الشعاعي والتـــي يشار إليها بالتكميم الشعاعـــي 2 (LVQ2) والتكميم الشعاعـــي 3 (LVQ3)، وسنتناول أولاً التكميم الشعاعي 2 (LVQ2) .

## 1.8.13 تعليم التكميم الشعاعي بمعلم

**Supervised learning vector quantization(LVQ)**

يشار إلى نماذج التكميم الشعاعي بمعلم بـــتعليم التكميم الشعاعي LVQ. الفرق الأساسي بين التكميم الشعاعي بدون معلم VQوتعليم التكميم الشعاعي LVQ بمعلم هو استعمال تصانيف خرج منشود معروفة $t(\mathbf{x}) = t$ لكل نموذج دخل $\mathbf{x}$.

ليكن C(x) هو صف x المحسوب بالشبكة. عندئذ، يوجد C(x) كما في حالة التكميم الشعاعي باستعمال :

$$\|\mathbf{w}_c - \mathbf{x}\| = \min_i \|\mathbf{w}_i - \mathbf{x}\|$$

عندما يكون الصف صحيحاً $(C(\mathbf{x}) = t)$ فإن شعاع الوزن للوحدة الرابحة يزاح باتجاه شعاع الدخل كما في حالة التكميم الشعاعي VQ. عندما يكون الأولي مختاراً بأسلوب غير صحيح $(C(\mathbf{x}) \neq t)$، فإن شعاع الوزن يزاح بعيداً عن شعاع الدخل. يمكن أن توصف قاعدة تحديث LVQ كما يلي:

$$\mathbf{W}_c^{new} = \mathbf{W}_c^{ald} + a(\mathbf{X} - \mathbf{W}_c) \qquad \text{في حالة } C(\mathbf{x}) = t$$

$$\mathbf{W}_c^{new} = \mathbf{W}_c^{ald} - a(\mathbf{X} - \mathbf{W}_c) \quad \text{في حالة } C(\mathbf{x}) \neq t \qquad (32.13)$$

$$\mathbf{W}_i^{new} = \mathbf{W}_i^{ald} \qquad \text{في حالة } \mathrm{i} \neq \mathrm{c}$$

سننظر في تغيرين لهذه القاعدة.

## 2.8.13 تعليم التكميم الشعاعي 2

**Learning vector quantization-2 (LVQ2)**

وظائف التعليم LVQ2 هي نفسها كما في حالة LVQ في تحديد صف شعاع الدخل (المعادلة (26.13). على أية حال، ينفذ التعليم فقط إذا تحققت الشروط التالية:

1. اختير صف شعاع الدخل بأسلوب غير صحيح $(C(\mathbf{x}) \neq t)$
2. شعاع الوزن الأولي الأقرب الثانـــي $\mathbf{W}_{C^*}$ هو صف صحيح
3. شعاع الدخل قريب للمستـــوى الفاصل بين شعاعي الوزن الأوليين المتجاورين أكثر $\mathbf{W}_C$ و $\mathbf{W}_{C^*}$.

عندما تتحقق هذه الشروط، تكون الأوزان المصاحبة للصف الصحيح $\mathbf{W}_{C^*}$ مزاحة باتجاه شعاع الدخل، وتكون الأوزان للصف غير الصحيح مزاحة بعيداً عن شعاع الدخل وفقاً لـــ:

$$\mathbf{W}_{c^*}^{new} = \mathbf{W}_{c^*}^{old} + a(\mathbf{X} - \mathrm{W}_{c^*})$$

$$\mathbf{W}_{c}^{new} = \mathbf{W}_{c}^{old} - a(\mathbf{X} - \mathrm{W}_{c}) \qquad (33.13)$$

$$\mathbf{W}_i^{new} = \mathbf{W}_i^{ald} \qquad \text{لكل } i \neq c, c^*$$

لقد ثبت أن هذه القاعدة تتمتع بخواص إنجاز جيدة (Kohonon عام 1988[179]).

لقد اقترحت تعديلات كبيرة على LVQ، أشير لها بـــ LVQ2.1 وLVQ3. وكلاهما تقوية محققة على تحديثات الأوزان للرابح ووحدة المكان الثانـــي عندما تقع نسبة المسافات

بين أشعة أوزان الدخل والرابح والدخل وأشعة أوزان الرابح الثانـــي تسقط ضمن النافذة الضيقة (Kohonen عام 1990[231][232])، التـــي تعرف كما يلي:

يقع شعاع الدخل x في النافذة إذا تحقق ما يلي:

$$\frac{d_c}{dr} > 1-\varepsilon \qquad (34.3)$$
$$\frac{d_r}{d_c} < 1+\varepsilon$$

حيث $d_c$ المسافة بين شعاع الدخل الحالي **x** والشعاع المرجع الأقرب إلى **x** (**yc**)، و $d_r$ المسافة بين **x** والشعاع المرجع الأقرب الثاني (**yr**) الذي يلي الشعاع المرجع الأقرب إلى **x**. حيث تعتمد قيمة $\varepsilon$ على عدد أمثلة التدريب؛ وتساوي قيمتها النموذجية 0.35 (Kohonen عام 1990[231]). في LVQ2، تكون الأشعة **yc** و**yr** محدثة إذا:

1.وقع شعاع الدخل **x** في النافذة

2. تنتمي الأشعة **yc** و**yr** إلى صفوف مختلفة

3. ينتمي **x** لنفس صف **yr**، وفق قاعدة التحديث التالية :

$$y_c(t+1) = y_c(t) - a(t)[\mathbf{x}(t) - y_c(t)] \qquad (35.13)$$
$$y_r(t+1) = y_r(t) + a(t)[\mathbf{x}(t) - y_r(t)]$$

## 3.8.13 تعليم التكميم الشعاعي 2.1 (LVQ2.1)

في التعديل المنفذ على تعليم التكميم الشعاعي والمعروف بـــ LVQ2.1، اعتبر Kohonen في مقالته–أ عام 1990[231] الشعاعين المرجعيين الأقربين **yc1** و**yc2**. لتحديث هذين الشعاعين يلزم واحد منهما، وليكن **yc1**، ينتمي إلى الصف الصحيح (في حالة شعاع الدخل الحالي **x**) على حين أن الآخر لا ينتمي إلى نفس الصف الذي ينتمي إليه **x**. وخلافاً للـــ LVQ، فإن LVQ2.1 لا يميز فيما إذا كان الشعاع الأقرب هو الممثل للصف الصحيح أو الصف غير الصحيح لدخل معطى. كذلك رأينا في LVQ2 أن **x** يجب أن يقع في النافذة حتـــى يتم التحديث. ويصبح الشرط اللازم تحققه للاختبار في حالة النافذة هو :

$$\min\left[\frac{d_{c1}}{d_{c2}},\frac{d_{c2}}{d_{c1}}\right]>1-\varepsilon$$
$$\max\left[\frac{d_{c1}}{d_{c2}},\frac{d_{c2}}{d_{c1}}\right]>1+\varepsilon \qquad (36.13)$$

في هذه الخوارزمية سينتج لدينا تعابير أكثر تعقيداً لأننا لا نعرف فيما إذا كان **x** أقرب إلى $y_{c1}$ أو إلى $y_{c2}$. إذا تحققت هذه الشروط، يحدث الشعاع المرجع الذي ينتمي إلى نفس الصف الذي ينتمي إليه **x** وفقاً لـــ:

$$y_{c1}(t+1)=y_{c1}(t)+a(t)[x(t)-y_{c1}(t)] \qquad (37.13)$$

ويحدث الشعاع المرجع الذي لا ينتمي إلى نفس الصف كـــ **x** وفقاً لـــ:

$$y_{c2}(t+1)=y_{c2}(t)-a(t)[x(t)-y_{c2}(t)] \qquad (38.13)$$

### 4.8.13 تعليم التكميم الشعاعي 3 (LVQ3)

هناك تقوية محققة وفقاً لـــمقالة- ب Kohonen عام 1990[232]، التـــي تسمح للشعاعين الأقربين أن يتعلما مادام شعاع الدخل يحقق شرط النافذة:

$$\min\left[\frac{d_{c1}}{d_{c2}},\frac{d_{c2}}{d_{c1}}\right]>(1-\varepsilon)(1+\varepsilon) \qquad (39.13)$$

حيث القيمة النموذجية لـــ $\varepsilon=0.2$. (لاحظ أن شرط النافذة هذا أيضاً جرى استعماله في LVQ2 في 1990Kohonen-b[232]).

إذا انتمى أحد الشعاعين الأقربين، وليكن $y_{c1}$، إلى نفس الصف الذي ينتمي إليه **x** وانتمى الشعاع الأقرب الثانـــي $y_{c2}$، إلى صف مختلف، فإن تحديث الأوزان سيكون كما في LVQ2.1. على أية حال، يوسع LVQ3 خوارزمية التدريب لتحقيق التدريب في حالة **x**، و $y_{c1}$، و$y_{c2}$ التـــي تنتمي إلى نفس الصف. في هذه الحالة، يكون تحديث الأوزان:

$$y_c(t+1)=y_c(t)+\beta/t)[x(t)-y_c(t)] \qquad (40.13)$$

لكل من $y_{c1}$ و$y_{c2}$.

سيكون معدل التعليم $\beta(t)$ ضعف$\alpha(t)$ المستعمل في حالة $y_{c1}$ و$y_{c2}$ عندما ينتميان إلى صفوف مختلفة. الضارب المناسب يكون عادة بين 0.1 و0.5 وبقيم أصغر موافقة للنافذة التي

هي أضيق:

$$\beta(t) = m\alpha(t) \quad \text{من أجل } 0.1 < m < 0.5 \quad (41.13)$$

هذا التعديل في عملية التعليم يؤكد أن الأوزان (الأشعة المرجعية) تستمر بالتقرب من توزيعات الصف وتمنع الأشعة المرجعية من التحرك بعيداً عن مكانها الأمثلي إذا استمر التعليم.

**مثال 1:**

سنشرح خوارزمية تعليم التكميم الشعاعي LVQ لخمسة أشعة مخصصة لصفين اثنين.

تمثل أشعة الدخل التالية الصفين 1 و2:

| الصف | الشعاع |
|---|---|
| 1 | (1, 1, 0, 0) |
| 2 | (0, 0, 0, 1) |
| 2 | (0, 0, 1, 1) |
| 1 | (1, 0, 0, 0) |
| 2 | (0, 1, 1, 0) |

أول شعاعين سيستعملان كقيم أولية للشعاعين المرجعيين. وهكذا، تمثل وحدة الخرج الأولى الصف 1، وتمثل وحدة الخرج الثاني الصف 2 (أي، $C_1=1$ و$C_2=2$). هذا سيدع أشعة الدخل الأخرى (0, 0, 1, 1), (1, 0, 0, 0), (0, 1, 1, 0) كأشعة تدريب. سنشرح تكراراً واحداً (دوراً واحداً) فقط وسنذكر الخوارزمية بوجه عام مع الحساب الموافق.

1. وضع القيم الأولية للأشعة المرجعية ومعدل التعليم $\alpha(0)$. $\mathbf{w}_1=(1, 1, 0, 0)$ $\mathbf{w}_2 = (0, 0, 0, 1)$ و $\alpha(0) = 0.1$.
2. مادام شرط التوقف غير محقق كرر الخطوات من 3 حتى 7
3. لكل شعاع دخل التدريب **x**، كرر الخطوات من 4 حتى 5.

في حالة شعاع الدخل $\mathbf{x} = (0, 0, 1, 1)$، و$T=2$ (الصف الصحيح أو الفئة لشعاع التدريب)، كرر الخطوات من 4 حتى 5

4. أوجد J بحيث يكون $\|\mathbf{X}-\mathbf{W}_J\|$ أصغرياً: $J=2$ باعتبار **x** أقرب إلى $\mathbf{w}_2$ منه إلى $\mathbf{w}_1$.
5. تحديث $\mathbf{W}_J$ كما يلي:

إذا كان $T = C_J$ فإن $\mathbf{W}_J^{new} = \mathbf{W}_J^{old} + \alpha(X - \mathbf{W}_J^{old})$

إذا كان $C_J \neq T$ فإن $\mathbf{W}_J^{new} = \mathbf{W}_J^{old} - \alpha(X - \mathbf{W}_J^{old})$

$C_j$ هو الصف أو الفئة الممثل بواسطة وحدة الخرج رقم j .

باعتبار T=2 و$C_2$=2، فإن تحديث $\mathbf{w}_2$ سيكون كما يلي:

$$\mathbf{w}_2 = (0, 0, 0, 1) + 0.1[(0, 0, 1, 1) - (0, 0, 0, 1)] = (0, 0, 0.1, 1)$$

3. في حالة شعاع الدخل $\mathbf{x} = (1, 0, 0, 0)$، و$T = 1$، كرر الخطوات من 4 حتـــى 5

4. J=1 باعتبار $\mathbf{x}$ أقرب إلى $\mathbf{w}_1$ منه إلى $\mathbf{w}_2$ .

5. باعتبار $T = 1$ و $C_1 = 1$، فإن تحديث $\mathbf{w}_1$ سيكون كما يلي:

$$\mathbf{w}_1 = (1, 1, 0, 0) + 0.1[(1, 0, 0, 0) - (1, 1, 0, 0)] = (0, 0.9, 0, 0)$$

3. في حالة شعاع الدخل $\mathbf{x} = (0, 1, 1, 0)$،و $T = 2$، كرر الخطوات من 4 حتـــى 5

4. J=1 باعتبار $\mathbf{x}$ أقرب إلى $\mathbf{w}_1$ منه إلى $\mathbf{w}_2$.

5. باعتبار T=2 و $C_1$=1، فإن تحديث $\mathbf{w}_1$ سيكون كما يلي:

$$\mathbf{w}_1 = (1, 0.9, 0, 0) - 0.1[(0, 1, 1, 0) - (1, 0.9, 0, 0)] = (1.1, 0.89 - 0.1, 0)$$

6. نهاية الدور الأول. خفض قيمة معدل التعليم

7. اختبر شرط التوقف (قد يكون عدد محدد من التكرارات أو الوصول إلى قيمة معدل تعليم صغير بقدر كافٍ).

**مثال هندسي 2:**

سنستعمل الآن LVQ لتمثيل نقاط في مربع واحدي وفقاً لانتمائها إلى واحد من الصفوف الأربعة. سيكون لدينا أربع وحدات تجمع، واحدة لكل صف. ستوضع الأوزان بقيم أولية بحيث تكون وحدات التجمع في بداية التدريب متوضعة في الزوايا الأربعة لمنطقة الدخل:

| الأوزان الأولية | | الصف |
|---|---|---|
| 0 | 0 | 1(U) |
| 0 | 1 | 2(O) |
| 1 | 1 | 3(Y) |
| 1 | 0 | 4(X) |

معطيات التدريب موضحة في الشكل (15.13)، ونتائج اختبار الشبكة على نفس نقاط الدخل المستعملة في التدريب موضحة في الأشكال (16.13)، (20.3)، (عادة لاتستخدم نفس المعطيات من أجل التدريب والاختبار معاً، ويجب اختبار الشبكة على معطيات مختافة من معطيات التدريب).

```
X X X X X Y Y Y Y
X X X X X Y Y Y Y
U U U U U U U Y Y
U U U U U U U Y Y
U U U U U U U O O
U U U U U U U O O
U U U U U U U O O
U U U U U U U O O
U U U U U U U O O
```

**الشكل 15.13**

معطيات التدريب

```
X X X X X Y Y Y Y
X X X X U Y Y Y Y
X X U U U U Y Y Y
X U U U U U U Y Y
U U U U U U U O O
U U U U U U U O O
U U U U U U O O O
U U U U U O O O O
U U U U U O O O O
```

**الشكل 16.13**

النتائج بعد دور واحد

```
X X X X X Y Y Y Y
X X X X X Y Y Y Y
X X U U U U Y Y Y
U U U U U U Y Y Y
U U U U U U U O O
U U U U U U O O O
U U U U U U O O O
U U U U U O O O O
U U U U U O O O O
```

**الشكل 17.13**

النتائج بعد دورين

الأوزان الأولية

| | | |
|---|---|---|
| "U" | 0.00 | 0.00 |
| "Y" | 1.00 | 1.00 |
| "O" | 1.00 | 1.00 |
| "X" | 0.00 | 1.00 |

الأوزان بعد دور واحد

| | | |
|---|---|---|
| "U" | 0.44 | 0.52 |
| "Y" | 0.90 | 0.93 |
| "O" | 1.03 | 0.17 |
| "X" | 0.13 | 1.02 |

الأوزان بعد دورين

| | | |
|---|---|---|
| "U" | 0.41 | 0.55 |
| "Y" | 0.88 | 0.92 |
| "O" | 1.03 | 0.24 |
| "X" | 0.22 | 1.02 |

```
X X X X X Y Y Y Y
X X X X X Y Y Y Y
X X U U U Y Y Y Y
U U U U U U Y Y Y
U U U U U U O O O
U U U U U U O O O
U U U U U U O O O
U U U U U O O O O
U U U U U O O O O
```

**الشكل 18.13**

النتائج بعد ثلاثة أدوار

```
X X X X X Y Y Y Y
X X X X X Y Y Y Y
X X X X X Y Y Y Y
U U U U U U Y Y Y
U U U U U U U O O
U U U U U U O O O
U U U U U U O O O
U U U U U U O O O
U U U U U U O O O
```

**الشكل 19.13**

النتائج بعد عشرة أدوار

```
X X X X X Y Y Y Y
X X X X X Y Y Y Y
X X X X X Y Y Y Y
U U U U U Y Y Y Y
U U U U U U O O O
U U U U U U O O O
U U U U U U O O O
U U U U U U O O O
U U U U U U O O O
```

**الشكل 20.13**

النتائج بعد أربعين دوراً

الأوزان بعد 3 أدوار

| | | |
|---|---|---|
| "U" | 0.36 | 0.57 |
| "Y" | 0.89 | 0.92 |
| "O" | 1.05 | 0.26 |
| "X" | 0.27 | 1.00 |

الأوزان بعد 10 أدوارً

| | | |
|---|---|---|
| "U" | 0.34 | 0.44 |
| "Y" | 0.89 | 0.91 |
| "O" | 1.10 | 0.26 |
| "X" | 0.30 | 1.03 |

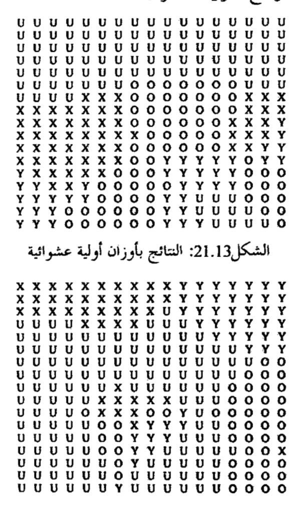
الأوزان بعد 40 دوراً

| | | |
|---|---|---|
| "U" | 0.30 | 0.31 |
| "Y" | 0.92 | 0.93 |
| "O" | 1.11 | 0.26 |
| "X" | 0.27 | 1.02 |

**مثال هندسي 3:**

لتحسين الأداء يجب أن نستعمل وحدات أكثر. باستعمال معطيات التدريب السابقة (مداخل التدريب للنقاط $(x, y)$: $x = 0.1\ i, (i=1,2,..,9)$; $y = 0.1\ j, (j=1,2,..,9)$ سنستعمل الآن 20 وحدة خرج، مع وضع أوزان أولية وتخصيص صفوف أولية عشوائياً. بالطبع هذا يتجاهل المعلومات المتوفرة من خلال عمليات اختيار القيم الأولية، لكن سنفعل ذلك بغية البرهان.

في الواقع العملي، يختار المرء عينة من النماذج التمثيلية من كل صف لتستعمل كأشعة مرجعية أولية. باستعمال معدل تعليم ثابت بقيمة 0.1، جرى تنفيذ 1000 دور تدريب. العدد الكبير اللازم كان نتيجة لوضع الأوزان عشوائياً.



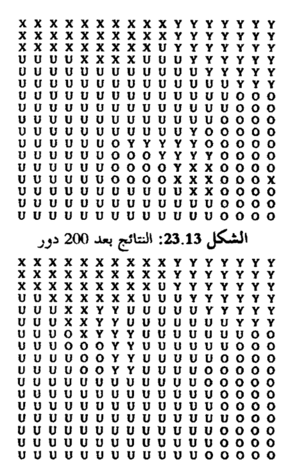

الشكل21.13: النتائج بأوزان أولية عشوائية



**الشكل 22.13**: النتائج بعد 100 دور

لتوضيح المناطق بأسلوب أفضل، سنختبر كل النقاط (x , y) في حالة:
((x = 0.05 i, (i=2,..,18); y = 0.05 j,(j=2,..,18)). تبين الأشكال (21.13)، (22.13) النتائج عند مراحل من التدريب. لاحظ أن معظم عمليات إعادة التوزيع لأشعة التجمعات تحدث خلال أول 100 دور. على أية حال، ثلاثة أشعة تجمعات (واحد من أجل كل من الصفوف 2 و3 و4) ستكون مقبوضة في المنطقة حيث تكون أشعة الدخل من الصف الأول (المرئية بالرمز "U"). تظهر النتائج بعد 100، 200، 400، 600، 800 دور أن الشبكة تزيح هذه الأشعة في اتجاهات متنوعة، قبل أن تدفع إلى الطرف الأيمن من الشكل.

التصنيف النهائي لنقاط الاختبار لم يحل ثانية منطقة بشكل L مرئية بالرمز "Y" مميزة بوضوح أكبر منها في المثال السابق. هذا وفقاً، على الأقل، للوضع العشوائي للأوزان. يحسن كثيراً وضع الأوزان الأولية وضعاً مناسباً لوحدات إنجاز LVQ .



```
X X X X X X X X X X Y Y Y Y Y Y Y
X X X X X X X X X X Y Y Y Y Y Y Y
X X X X X X X X X U Y Y Y Y Y Y Y
U U U U X X X X U U U Y Y Y Y Y Y
U U U U U U U U U U U U Y Y Y Y Y
U U U U U U U U U U U U U U Y Y Y
U U U U U U U U U U U U U U O O O
U U U U U U U U U U U U U U O O O
U U U U U U U U U U U U U O O O O
U U U U U U U U U U U Y O O O O O
U U U U U U O Y Y Y Y Y O O O O O
U U U U U U O O O Y Y Y Y O O O O
U U U U U U O O O O Y X X O O O O
U U U U U U O O O O X X X O O O X
U U U U U U U U U U U X X O O O O
U U U U U U U U U U U U U O O O O
U U U U U U U U U U U U U O O O O
```

**الشكل 23.13**: النتائج بعد 200 دور

```
X X X X X X X X X X Y Y Y Y Y Y Y
X X X X X X X X X X Y Y Y Y Y Y Y
X X X X X X X X X U Y Y Y Y Y Y Y
U U X X X X X X U U U Y Y Y Y Y Y
U U U X X Y Y U U U U U Y Y Y Y Y
U U U X X Y Y U U U U U U U Y Y Y
U U U O X Y Y Y U U U U U U U O O
U U U O O O Y Y U U U U U U O O O
U U U U O O Y Y U U U U U O O O O
U U U U O O Y Y U U U U O O O O O
U U U U U U U U U U U U O O O O O
U U U U U U U U U U U U U O O O O
U U U U U U U U U U U U U O O O O
U U U U U U U U U U U U U U O O O
U U U U U U U U U U U U U O O O O
U U U U U U U U U U U U U O O O O
U U U U U U U U U U U U O O O O O
```

**الشكل 24.13**: النتائج بعد 400 دور



**الشكل25.13**: النتائج بعد 600دور

**الشكل 26.13**: النتائج بعد 800 دور

**الشكل 27.13**: النتائج بعد 1000 دور

## 9.13 شبكات خريطة الملامح الذاتية التنظيم

## Self-Organizing Feature Map networks(SOFM)

ربما تكون القشرة الدماغية لمخ الإنسان أكثر الأنظمة البيولوجية تعقيداً. فهي على المستوى المصغر (micro level)، منظمة في طبقات عديدة من العصبونات بكثافات وأنواع مختلفة، وعلى المستوى المكبر (macro level) منظمة في مناطق مكانية وفقاً لوظيفة جسمية معينة. مثلاً، هناك منطقة الرؤية، ومنطقة حركة العين، والسمع، والكلام، واللمس، والتفكير،...الخ. تتألف كل منطقة من عدد ضخم من العصبونات المتشابهة التـــي تتعاون عندما تنفذ وظائف خاصة، وتصبح مختصة عند المعالجة. توافق كل طبقة من الطبقات تطبيق مجموعة وظيفية ما لمداخل الحواس، مثل القشرة البصرية، ومستقبلات السمع، ووظائف الحركة، وقشرة اللمس، والتفكير، ...الخ.

تستجيب مجموعات العصبونات التـــي تقع ضمن كل منطقة بطريقة مشتركة للإثارات من الخلايا الحسية الفعالة. مثلاً، تستجيب العصبونات في القشرة البصرية لنماذج ضوئية معينة تسقط على الشبكية، وتصبح خلايا منطقة قشرة اللمس مهيجة بالمداخل من الخلايا الحسية تحت الجلد، وتستجيب خلايا الخريطة السمعية في مجموعات متوضعة لأصوات مختلفة بنيت على التردد أو النغمة.

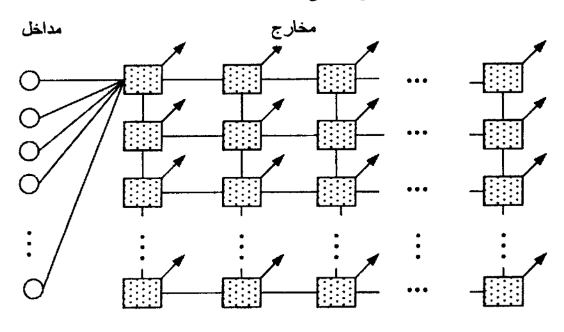
تكون حقول استقبال هذه العصبونات المنظمة مكانياً مرتبطة مباشرة مع العصبونات الحسية. هناك تطبيق (mapping) أو إسقاط للملامح من العصبونات الحسية إلى المناطق المكانية المرافقة أو القشرة. جرت نمذجة تطبيق الملمح البيولوجي هذا للدماغ بوجه مسؤول بالشبكات العصبونية الصنعية على شكل خارطة ذاتية التنظيم.

إن الشبكة العصبونية الصنعية على شكل خارطة ذاتية التنظيم هي نموذج مبسط لإسقاط أو تطبيق ملمح إلى منطقة متوضعة في الدماغ، والتـــي اشتق اسمها من هذا النموذج البيولوجي، هذه الشبكة تسمى بخارطة المحافظة الطبولوجية؛ حيث يفرض وجود بنية طبولوجية فيما بين الوحدات. وهي شبكة ذاتية التنظيم، تنافسية، تتعلم من الوسط المحيط دون مساعدة من معلم.

بنية الشبكة بسيطة جداً، فهي تتألف من مجموعة من العصبونات المنظمة هندسياً في بعد أو بعدين أو ثلاثة أو أكثر. فالشبكة أحادية البعد هي طبقة وحيدة من الوحدات المرتبة على شكل سطر.

وفي حالة الشبكة ثنائية البعد، تكون الوحدات مرتبة كمصفوفة تصالبية، وهكذا في أبعاد أكثر.

التصالب ثنائي البعد للوحدات موضح في الشكل (28.13). الوصلات الواضحة في الشكل هي المداخل والمخارج والوصلات المتجاورة مباشرة. أما الوصلات الداخلية للوحدات المتباعدة فقد حذفت من الشكل لتبسيطه.




شبكية ثنائية البعد من العصبونات

**الشكل28.13**: شبكة خريطة الملامح الذاتية التنظيم الثنائية البعد

يكون شعاع ملامح الدخل **x** ببعد n موصلاً توصيلاً كاملاً لكل وحدات الشبكة من خلال أشعة وزن متكيفة $\mathbf{w}_j$، $j = 1,2,...,m$. كما في حالة شبكات VQ، تربح المنافسة الوحدة ذات شعاع الوزن الأقرب لنموذج الدخل. يستجيب الرابح أعظمياً ويقود كل الوحدات الأخرى إلى خرج صفري.

يؤدى فعل المنافسة من خلال وصلات (وزن ثابت) جانبية بين الوحدات المتجاورة،

حيث تكون التهيجات والتخميدات متولدة، وخلافاً لشبكاتVQ، ليست الوحدة الرابحة فقط هي المستفيدة من التعليم التالي للمنافسة. تشارك الوحدة الرابحة تجربة التعليم مع أقرب جاراتها، وتنفذ عملية التعليم بطريقة ما بحيث تميل العناصر المتجاورة إلى اصطفاف أوزانها في نفس الاتجاه كنموذج الدخل، على حين تنتظم أوزان الوحدات التي هي أبعد في الاتجاهات المتعاكسة.

تعرف مناطق الجوار للوحدة J بقيمة أنصاف الأقطار R، مثلاً، لسطر من الوحدات m، يتألف الجوار بنصف القطر R حول الوحدة J من كل الوحدات j التـــي تحقق المتراجحة:

$$\max(1, J - R) \le j \le \min(J + R, m)$$

كما هو موضح في الشكل (29.13) في حالة m = 10.

```
*   *   *   {*   (*   [#j]   *)   *}   *   *   *
1                                              m
            {}R = 2     ()R = 1     []R = 0
```

**الشكل 29.13**: مصفوفة سطر من الوحدات

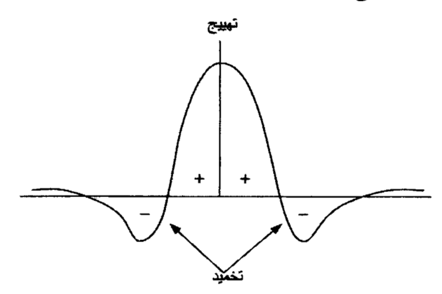
في الفقرات الأخيرة، سنركز على بنية خريطة الملامح الذاتية التنظيم الثنائية البعد. من أجل هذا، من المناسب الإشارة لكل وحدة بشعاع **r** مكان إحداثياته (x,y). وهكذا، فإن دخل الوحدة رقم **r** هو تهييج خارجي $\mathbf{x}^T\mathbf{w_r}$ إضافة إلى تغذية عكسية داخلية من الوحدات $\mathbf{r}'$ المتصلة مع الوحدة **r** من خلال الوزان المثبتة $\mathbf{v_{rr'}}$. الدخل من الوحدات الداخلية المتصلة مع الوحدة **r** يعطى بواسطة $\mathbf{y_{r'}}^T\mathbf{v_{rr'}}$. يمكن أيضاً أن تملك الوحدات عتبة تهييج $\theta$. يعطى تفعيل الخرج الوحيد من الوحدة **r** بواسطة $\mathbf{y_r}$ كما يلي:

$$y_r = f\left(\sum_i W_{ri} xi + \sum_{r'} V_{rr'} y_{r'} - \theta\right) \qquad (42.13)$$

مجال المجاميع في المعادلة (*42.13*) يشمل أدلة كل وصلات الدخل والوحدات الداخلية المتصلة مع الوحدة **r**، والتابع **f** تابع تفعيل غير خطي ما، مثل **sigmoid**.

يمكن وصف الديناميكيات الكاملة لخريطة الملامح الذاتية التنظيم بواسطة معادلات تفاضلية مشابهة لتلك المعطاة سابقاً في شبكات VQ (المعادلات (24.13) و (25.13)) التـــي تأخذ بالحسبان مقياس المتحولات وتوابع تفعيل غير خطية f للوحدات. تسلك منبهات

التعليم المتولدة بواسطة الوحدة الرابحة تهييج "مركز فعال ومحيط غير فعال" يشبه شكل القبعة المكسيكية كما هو موضح في الشكل (30.13).




**الشكل 30.13**: شدات التفاعل الجانبية المشابه لشكل القبعة المكسيكية

الوحدة القريبة للوحدة الرابحة تهيج أكثر من الوحدات الأكثر بعداً، والوحدات البعيدة نوعاً ما تكون ممنوعة، أي أوزانها تزاح بعيداً عن اتجاه شعاع الدخل. بعد أن يتقدم التعليم لبعض الوقت، تميل أشعة الوزن في مصفوفة الوحدات المجمعة إلى نموذج توزيع احتمال نماذج الدخل من خلال خريطة المحافظة على الملامح طبولوجياً.

كما في حالة شبكات VQ، الشكل البسط لعمل شبكة خريطة الملامح الذاتية التنظيم يمكن أن يستعمل لتقريب ديناميكيات النظام. بافتراض أن نماذج الدخل كلها معيارية بطول واحدي وأن أشعة وزن الدخل بطول ثابت، $\|\mathbf{w}_\mathbf{r}\| = c$، نستطيع كتابة معادلات التفعيل المبسطة كما يلي:

$$\|\mathbf{w}_\mathbf{r} - \mathbf{x}\| = \min_{\mathbf{r}'} \|\mathbf{w}_{\mathbf{r}'} - \mathbf{x}\| \qquad (43.13)$$

حيث الوحدة **r** هي الرابحة للمنافسة وشعاع وزنها يكون الأقرب إلى نموذج الدخل **x**.

يستمر التعليم بعدئذ وفقاً للقاعدة المعطاة كما يلي:

$$\mathbf{W}_\mathbf{r}^{\text{new}} = \mathbf{W}_\mathbf{r}^{\text{old}} + \alpha \mathbf{h}_{\mathbf{rr}'} (\mathbf{X} - \mathbf{W}_\mathbf{r}^{\text{old}}) \qquad (44.13)$$

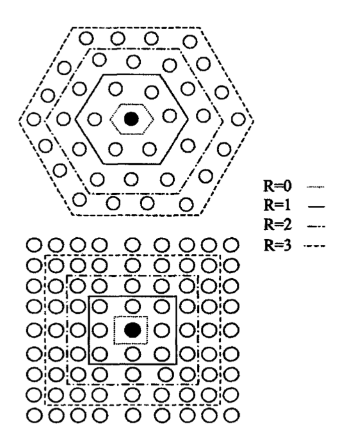
حيث $\mathbf{h}_{\mathbf{rr}'}$ تابع الجوار بقيمة أعظمية مركزة عند الوحدة الرابحة **r** (الوحدة الغامقة ذات

نصف قطر الجوار R المساوي إلى الصفر R= 0 في الشكل (31.13)) ويصبح صفراً كلما ازدادت المسافة بين r والوحدات المجاورة `r، أي $h_{rr`}$ تعرف في حدود المسافة بين r و `r.

نستعمل $N_r$ للدلالة على جوار الوحدة r بحيث $h_{rr`} > 0$ لكل `r ضمن الجوار.

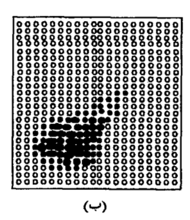
عامل التعليم الموجب α هو تابع لخطوة التعليم ويتناقص باتجاه الصفر كلما تقدم التعليم. وبغية التبسيط، يؤخذ شكل $h_{rr`}$ أحياناً ليكون تابعاً، بشكل قمة مسطحة، للمسافة عبر منطقة هندسية مثل الشكل المربع أو السداسي كما هو موضح في الشكل (31.13).

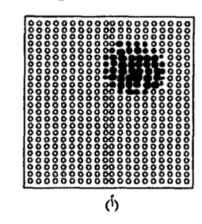
لكل الوحدات ضمن الجوار أوزان معدلة في اتجاه نموذج الدخل، ولا تستقبل الوحدات في الخارج أي تعديل. كلما قدمت نماذج دخل أكثر فأكثر إلى الشبكة، ينقص حجم الجوار حتـــى يشمل فقط الوحدة الرابحة وبعض الجوار الأقرب. مبدئياً، تكون قيم الأوزان عشوائية كثيراً أو قليلاً.




**الشكل 31.13**: مناطق الجوار المتكيفة المربعة والسداسية للوحدة الرابحة (الغامقة)

كلما تقدم التعليم، تصبح مناطق شكل الفعالية عبر الوحدة الرابحة ذات شكل يشبه الفقاعات. محاكاة تابع التفعيل عبر الوحدة الرابحة موضحة في الشكل (32.13 أ) حيث تتشكل فقاعات التفعيل عبر شبكة ثنائية البعد خلال التعليم. تظهر في الشكل (32.13ب) المحاكاة كيف تنـــزاح منطقة الفعالية كلما عانت نماذج الدخل من إزاحة في المكان. الدوائر السوداء هي وحدات بفعالية مزادة في المصفوفة وحجم الدوائر يشير إلى مستوى الفعالية. يظهر الذيل على الطرف الأيمن للشكل (32.13 ب) اتجاه الحركة في نماذج الدخل.


(ب)

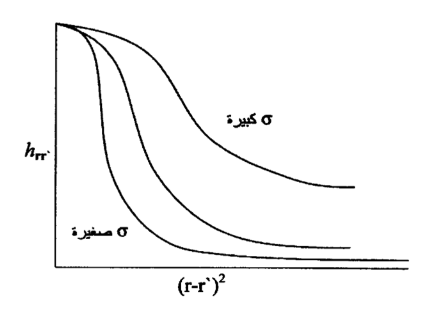

(أ)

**الشكل 32.13**: فقاعات الفعالية عبر الشبكة الثنائية البعد (أ) منطقة فعالية ثابتة، (ب) الدخل يتغير ببطء مع فقاعة متحركة

عوضاً عن استعمال حدود جوار حادة لـــ $\mathbf{N_r}$، يمكن أن نختار $\mathbf{h_{rr'}}$ ليكون تابع غوص ناعماً من الشكل:

$$h_{rr'} = \exp[-(\mathbf{r}.\mathbf{r}')^2/(2\sigma^2)]) \qquad (45.13)$$

حيث يعرف σ نصف قطر الجوار. مبدئياً، يختار الوسيط σ بقيمة كبيرة بحيث يشترك عدد ضخم من الوحدات المتجاورة في تجربة التعليم مع الرابح. كلما ازداد عدد خطوات التعليم، فإن قيمة σ تتناقص تدريجياً إلى قيمة ثابتة صغيرة تجعل منطقة الجوار أكثر انتقائية. إن شكل تابع الجوار الغوصي لقيم مختلفة لـــ σ موضح في الشكل (33.13). عندما تتناقص قيمة σ تميل الأوزان إلى التقارب وشكل الصورة الطبولوجية إلى فراغ الدخل.




**الشكل 33.13:** الجوار الغوصي كتابع لـــ σ

يمكن تلخيص خوارزمية تعليم شبكة خريطة الملامح الذاتية التنظيم كما يلي:

1. وضع القيم الأولية للأوزان $\mathbf{w_r}$ بمتوسطات مناسبة (أعداد عشوائية صغيرة)، ووضع وسطاء معدل التعلم ووسطاء الجوار الطبولوجي. كرر الخطوات من 2 إلى 6 حتـــى تستقر أوزان الشبكة
2. اختيار الشعاع $\mathbf{x}$ من توزيع نموذج الدخل لدخل الشبكة، ولكل شعاع دخل كرر الخطوات من 3 إلى 4.
3. تحديد وحدة المصفوفة $\mathbf{r}$ مع شعاع وزن أقرب إلى $\mathbf{x}$ بحساب:

$$\|\mathbf{w}_{\mathbf{r}} - \mathbf{x}\| = \min_{\mathbf{r'}} \|\mathbf{w}_{\mathbf{r'}} - \mathbf{x}\|$$

4. تحديث أشعة الوزن على التكرار رقم t+1 وفقاً للمعادلة:

للوحدات $\mathbf{N_r} \in \mathbf{r}$ $\quad \mathbf{W}_r(t+1) = \mathbf{W}_r(t) + \alpha(t)\, h_{rr'}\, (\mathbf{X} - \mathbf{W}_r(t))$

للوحدات $\mathbf{N_r} \notin \mathbf{r}$ $\quad \mathbf{w_r}(t+1) = \mathbf{w_r}(t)$

حيث $\mathbf{N_r}$ هو جوار $\mathbf{r}$ كما ذكرنا من قبل.

5. تقليل وسطاء الجوار ومعدل التعليم.
6. اختبار شرط التوقف.

لاحظ أنه إذا قللت $\mathbf{N_r}$ لتشمل الوحدة الرابحة فقط، عندئذ تنجزخريطة الملامح الذاتية

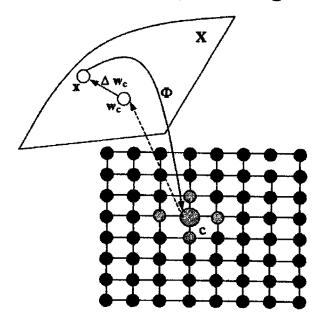
التنظيم تكميماً شعاعياً، وتصبح بصفة أساسية شبكة VQ كما وصف في الفقرة (7.13)، وتكون مناطق القرار مشكلة بواسطة الأوزان المتكيفة $\mathbf{w_r}$ مع حدود معرفة بواسطة مضلعات Voronoi. كل نقطة ضمن المنطقة المعطاة تكون أقرب إلى شعاع الوزن المرجع $\mathbf{w_r}$ لتلك المنطقة من أي شعاع وزن آخر، وتكون حدود مستويات فصل المناطق متعامدة مع خطوط وصل أشعة وزن المنطقة المحاورة.

بعد أن يتقدم التعليم بقدر كاف بحيث تستقر الأوزان، تنجز خريطة الملامح الذاتية التنظيم المدربة تطبيقاً من الجملة المولِّدة X لفراغ نموذج الدخل إلى مركز التهييج c في الشبكة الذي يبدو مشابهاً لتطبيق مستمر للدخل المشكل عبر الشبكة كالمعين بواسطة القيم النسبية لأوزان الشبكة.

يعتمد محل مركز التهييج الأعظمي على اتجاه شعاع الدخل **x** الذي طبق إلى مكان **r** في المصفوفة الثنائية البعد.

بالسماح للوسيط $\alpha(t)$ بالتناقص مع التعليم، بشرط أن يبقى موجباً، يمكن أن تحتفظ الشبكة بلدونتها (مطاوعتها)، وتستمر بالتكيف مع التغيرات في الوسط المحيط.

يوضح الشكل (34.13) تطبيق نموذج إلى فراغ وزن، حيث **Φ** هو التطبيق من **X** إلى مصفوفة شبكة خريطة الملامح الذاتية التنظيم.




**الشكل 34.13**: تطبيق نموذج إلى فراغ

لتوضيح عمل الشبكة وخوارزميتها سنقوم بمناقشة بعض الأمثلة البسيطة.

**مثال 4:**

لتكن لدينا أشعة الدخل التالية، والمطلوب تجميعها في m تجمعاً:

(1, 1, 0, 0), (0, 0, 0, 1), (1, 0, 0, 0), (0, 0, 1, 1)

العدد الأعظمي للتجمعات الممكن تشكيلها هو $m = 2$. لنفترض أن معدل التعليم (المتناقص) هو: $\alpha(0) = 0.6$ و$\alpha(t+1) = 0.5\alpha(t)$. بتجمعين متوفرين فقط، يوضع جوار العقدة J بحيث يحدث تجمع واحد فقط أوزانه عند كل خطوة (هذا يعني $R = 0$). لزيادة الإيضاح سنقوم بعرض الخوارزمية بوجه عام مع التطبيق الحسابي المقابل.

1. وضع القيم الأولية للأوزان $w_{ij}$، ووسطاء معدل التعليم، والجوار الطبولوجي، مصفوفة الأوزان الأولية هي:

$$\begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.5 & 0.7 \\ 0.9 & 0.3 \end{bmatrix}$$

نصف قطر الجوار الأولي: $R = 0$، ومعدل التعليم الأولي: $\alpha(0) = 0.6$.

2. بداية التدريب. مادام شرط التوقف غير محقق كرر مايلي:
3. لكل شعاع دخل x ،كرر الخطوات من 4 إلى 6.

   في حالة شعاع الدخل الأول (1,1,0,0).
4. لكل j، احسب (مربع المسافة الإقليدية):

$$D(j) = \sum_i (w_{ij} - x_i)^2$$

$$D(1) = (0.2-1)^2 + (0.6-1)^2 + (0.5-0)^2 + (0.9-0)^2 = 1.86$$

$$D(2) = (0.8-1)^2 + (0.4-1)^2 + (0.7-0)^2 + (0.3-0)^2 = \underline{\mathbf{0.98}}$$

5. إيجاد الدليل J بحيث تكون قيمة D أصغرية؛ سيكون شعاع الدخل أقرب لعقدة الخرج رقم 2، ومن ثم: $J = 2$.
6. لكل الوحدات j ضمن الجوار المخصص J، ولكل $i$، حدث الأوزان وفقاً لـ:

$$w_{ij}^{new} = w_{ij}^{old} + \alpha\left[x_i - w_{ij}^{old}\right]$$

ستكون الأوزان على الوحدة الرابحة:

$$w_{i2}^{new} = w_{i2}^{old} - 0.6[x_i - w_{i2}^{old}]$$
$$= 0.4w_{i2}^{old} + 0.6x_i$$

هذا يعطي مصفوفة الأوزان (تحديث العمود الثاني) التالية:

$$\begin{bmatrix} 0.2 & 0.92 \\ 0.6 & 0.76 \\ 0.5 & 0.28 \\ 0.9 & 0.12 \end{bmatrix}$$

3. في حالة شعاع الدخل الثاني، (0,0,0,1) كرر الخطوات من 4 إلى 6

4.

$$D(1) = (0.2 - 0)^2 + (0.6 - 0)^2 + (0.5 - 0)^2 + (0.9 - 1)^2 = \underline{\mathbf{0.66}}$$

$$D(2) = (0.92 - 0)^2 + (0.76\text{-}0)^2 + (0.28 - 0)^2 + (0.12 - 1)^2 = 2.2768$$

5. سيكون شعاع الدخل أقرب لعقدة الخرج رقم 1، ومنه: J=1 .

7. ستكون الأوزان على الوحدة الرابحة محدثة وفقاً لـــ:

$$w_{i1}^{new} = w_{i1}^{old} - 0.6[x_i - w_{i1}^{old}]$$
$$= 0.4w_{i1}^{old} + 0.6x_i$$

هذا يعطي مصفوفة وزن :

$$\begin{bmatrix} 0.08 & 0.92 \\ 0.24 & 0.76 \\ 0.20 & 0.28 \\ 0.96 & 0.12 \end{bmatrix}$$

3. في حالة شعاع الدخل الثالث، (1,0,0,0) كرر الخطوات من 4 إلى 6

4.

$$D(1) = (0.08-1)^2 + (0.24\text{-}0)^2 + (0.2\text{-}0)^2 + (0.96 - 0)^2 = 1.8656$$

$$D(2) = (0.92-1)^2 + (0.76\text{-}0)^2 + (0.28\text{-}0)^2 + (0.12 - 1)^2 = \underline{\mathbf{0.6768}}$$

5. سيكون شعاع الدخل أقرب لعقدة الخرج رقم 2، ومنه: J = 2.

6. الأوزان على الوحدة الرابحة ستكون محدثة وفقاً لـــ:

$$w_{i2}^{new} = w_{i2}^{old} - 0.6[x_i - w_{i2}^{old}]$$
$$= 0.4w_{i2}^{old} + 0.6x_i$$

وهذا يعطي مصفوفة وزن:

$$\begin{bmatrix} 0.08 & 0.968 \\ 0.24 & 0.304 \\ 0.20 & 0.112 \\ 0.96 & 0.048 \end{bmatrix}$$

3. في حالة شعاع الدخل الرابع، (0, 0, 1, 1) كرر الخطوات من 4 إلى 6

.4

$$D(1) = (0.08-0)^2 + (0.24 - 0)^2 + (0.2-1)^2 + (0.96-1)^2 = \underline{\mathbf{0.7056}}$$
$$D(2) = (0.968-0)^2 + (0.304 - 0)^2 + (0.112-1)^2 + (0.048-1)^2 = 2.724$$

5. سيكون شعاع الدخل أقرب لعقدة الخرج رقم 1، ومنه: J=1.

6. ستكون الأوزان على الوحدة الرابحة محدثة وفقاً لـــ :

$$w_{i1}^{new} = w_{i1}^{old} - 0.6[x_i - w_{i1}^{old}]$$
$$= 0.4w_{i1}^{old} + 0.6x_i$$

وهذا يعطي مصفوفة وزن :

$$\begin{bmatrix} 0.032 & 0.968 \\ 0.096 & 0.304 \\ 0.680 & 0.112 \\ 0.984 & 0.048 \end{bmatrix}$$

7.إنقاص معدل التعليم

معادلات تحديث الوزن أصبحت الآن:

$$w_{ij}^{new} = w_{ij}^{old} + 0.3[x_i - w_{ij}^{old}]$$
$$= 0.7w_{ij}^{old} + 0.3x_i$$

ستكون مصفوفة الوزن بعد الدور الثانـــي من التدريب كما يلي:

$$\begin{bmatrix} 0.016 & 0.980 \\ 0.047 & 0.360 \\ 0.630 & 0.055 \\ 0.999 & 0.024 \end{bmatrix}$$

وبإجراء إنقاص خطي لمعدل التعليم متكرر من 0.6 إلى 0.01 عبر 100 دور حصلنا على النتائج التالية:

الدور الصفري: مصفوفة الوزن هي:

$$\begin{bmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.5 & 0.7 \\ 0.9 & 0.3 \end{bmatrix}$$

الدور الأول: مصفوفة الوزن هي:

$$\begin{bmatrix} 0.032 & 0.970 \\ 0.096 & 0.300 \\ 0.680 & 0.110 \\ 0.980 & 0.048 \end{bmatrix}$$

الدور الثاني: مصفوفة الوزن هي:

$$\begin{bmatrix} 0.0053 & 0.9900 \\ -0.1700 & 0.3000 \\ 0.7000 & 0.0200 \\ 1.0000 & 0.0086 \end{bmatrix}$$

الدور العاشر:مصفوفة الوزن هي:

$$\begin{bmatrix} 1.5e-7 & 1.0000 \\ 4.6e-7 & 0.3700 \\ 0.6300 & 5.4e-7 \\ 1.0000 & 2.3e-7 \end{bmatrix}$$

الدور الخمسين: مصفوفة الوزن هي:

$$\begin{bmatrix} 1.9e-19 & 1.0000 \\ 5.7e-15 & 0.4700 \\ 0.5300 & 6.6e-15 \\ 1.0000 & 2.8e-15 \end{bmatrix}$$

الدور المئة: مصفوفة الوزن هي:

$$\begin{bmatrix} 6.7e-17 & 1.0000 \\ 2.0e-16 & 0.4900 \\ 0.5100 & 2.3e-16 \\ 1.0000 & 1.0e-16 \end{bmatrix}$$

يظهر أن مصفوفات الأوزان تتقارب إلى المصفوفة:

$$\begin{bmatrix} 0.0 & 1.0 \\ 0.0 & 0.5 \\ 0.5 & 0.0 \\ 1.0 & 0.0 \end{bmatrix}$$

العمود الأول هو متوسط الشعاعين المتوضعين في التجمع الأول، والعمود الثاني هو متوسط الشعاعين المتوضعين في التجمع الثاني.

**مثال 5:**

استخدمت شبكة خريطة الملامح الذاتية التنظيم في مسائل تعرف الأشكال، سنوضح بواسطة المثال الحالي كيف استخدمت هذه الشبكة لتشكيل تجمعات نماذج الدخل الممثلة بسبعة أحرف من ثلاثة تشكيلات مختلفة. نماذج الدخل من التشكيلات الثلاثة موضحة في الشكل (35.13).

لمعرفة تأثير البنية الطبولوجية التي ستفترض بين الوحدات (كما شرحنا في النص من قبل) سنحاول حل هذه المسألة بثلاث حالات: بدون بنية وببنية خطية وببنية مستطيلة، وسنستخدم 25 وحدة، هذا يعني أنه يمكن تشكيل 25 تجمعاً.

**إذا لم يفرض بنية معينة لوحدات التجمعات؛ فهذا يعني، السماح للوحدة الرابحة فقط بتعلم النموذج المقدم** (أي $R = 0$)، وسنحصل على 21 نموذجاً تشكل 5 تجمعات.

الدخل من التشكيلة الأولى :

```
00##000  ######0  00#####  #####00  #######  000####  ###00##
000#000  0#0000#  0#0000#  0#000#0  0#0000#  00000#0  0#00#00
000#000  0#0000#  #000000  0#0000#  0#00000  00000#0  0#0#000
00#0#00  0#0000#  #000000  0#0000#  0#0#000  00000#0  0##0000
00#0#00  0#####0  #000000  0#0000#  0###000  00000#0  0##0000
0#####0  0#0000#  #000000  0#0000#  0#0#000  00000#0  0#0#000
0#000#0  0#0000#  #000000  0#0000#  0#00000  0#000#0  0#00#00
0#000#0  0#0000#  0#0000#  0#000#0  0#0000#  0#000#0  0#000#0
###0###  ######0  00####0  #####00  #######  00###00  ###00##
A......  .B.....  ..C....  ...D...  ....E..  .....J.  ......K
```

**الشكل (35.13)**(أ): نماذج دخل التدريب

الدخل من التشكيلة الثانية

```
000#000  ######0  00###00  #####00  #######  00000#0  #0000#0
000#000  #00000#  0#000#0  #0000#0  #000000  00000#0  #000#00
000#000  #00000#  #00000#  #00000#  #000000  00000#0  #00#000
00#0#00  #00000#  #000000  #00000#  #000000  00000#0  #0#0000
00#0#00  ######0  #000000  #00000#  #####00  00000#0  ##00000
0#000#0  #00000#  #000000  #00000#  #000000  00000#0  #0#0000
0#####0  #00000#  #00000#  #00000#  #000000  0#000#0  #00#000
0#000#0  #00000#  0#000#0  #0000#0  #000000  0#000#0  #000#00
0#000#0  ######0  00###00  #####00  #######  00###00  #0000#0
A......  .B.....  ..C....  ...D...  ....E..  .....J.  ......K
```

الدخل من التشكيلة الثالثة :

```
000#000  ######0  00###0#  #####00  #######  0000###  ###00##
000#000  0#0000#  0#000##  0#000#0  0#0000#  00000#0  0#000#0
00#0#00  0#0000#  #00000#  0#0000#  0#00#00  00000#0  0#00#00
00#0#00  0#0000#  #000000  0#0000#  0####00  00000#0  0#0#000
0#000#0  0#####0  #000000  0#0000#  0#00#00  00000#0  0##0000
0#####0  0#0000#  #000000  0#0000#  0#00000  00000#0  0#0#000
#00000#  0#0000#  #00000#  0#0000#  0#00000  00000#0  0#00#00
#00000#  0#0000#  0#000#0  0#000#0  0#0000#  0#000#0  0#000#0
##000##  ######0  00###00  #####00  #######  00###00  ###00##
A......  .B.....  ..C....  ...D...  ....E..  ....J.  ......K
```

تتمة **الشكل (35.13)**(ب): نماذج دخل التدريب

| الوحدة | النماذج | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | | | | | $C_1$ | $C_2$ | $C_3$ |
| 13 | $B_1$ | $B_3$ | $D_1$ | $D_3$ | $E_1$ | $K_1$ | $K_3$ |
| 16 | | | | | $A_1$ | $A_2$ | $A_3$ |
| 18 | | | | | $J_1$ | $J_2$ | $J_3$ |
| 24 | | | | $B_2$ | $D_2$ | $E_2$ | $K_2$ |

**مثال 6:**

أما في حالة البنية الطبولوجية الخطية (R = 1) فسنحصل على توزيع أفضل للنماذج على وحدات التجمعات المتوفرة. العقدة الرابحة J وجاراها الطبولوجية (J+1 و J–1) سيسمح لها بالتعلم في كل دور.

لاحظ أن عقد الجوار التـــي تتعلم أيضاً ليس لها مبدئياً أشعة وزن قريبة من نموذج الدخل.

| الوحدة | النماذج | | | |
|---|---|---|---|---|
| 6 | | | | $K_2$ |
| 10 | | $J_1$ | $J_2$ | $J_3$ |
| 14 | | | $E_1$ | $E_3$ |
| 16 | | | $K_1$ | $K_3$ |
| 18 | $B_1$ | $B_3$ | $D_1$ | $D_3$ |
| 20 | | $C_1$ | $C_2$ | $C_3$ |
| 22 | | | | $D_2$ |
| 23 | | | $B_2$ | $E_2$ |
| 25 | | $A_1$ | $A_2$ | $A_3$ |

لاحظ أيضاً أنه في حالات عديدة هناك وحدات غير مستعملة بين زوج الوحدات التـــي لها تجمعات نماذج مصاحبة لها. هذا يقترح أن الوحدات التـــي سحبت باتجاهات متعاكسة خلال التدريب لن تتعلم أي نموذج (بكلمات أخرى، في معظم الحالات، نماذج الدخل هذه تشكل صفوفاً مختلفة جداً).

**مثال 7:**

أما في حالة الطبولوجية الثنائية البعد المفروضة على الوحدات، سيكون لكل وحدة دليلان. إذا كانت الوحدة $X_{IJ}$ هي الرابحة، فإن جاراتها $X_{I+1,J}, X_{I-1,J}, X_{I,J+1}, X_{I,J-1}$ أيضاً ستتعلم. وهذا يعطي شكلاً طبولوجياً معيناً (ليس مستطيلاً) عوضاً عن الشكل المستطيل الكامل الموضح في الشكل(31.13).

النتائج موضحة في الشكل (36.13) كما يلي:

| i/j | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | | $J_1,J_2,J_3$ | | $D_2$ | |
| 2 | $C_1,C_2,C_3$ | | $D_1,D_3$ | | $B_2,E_2$ |
| 3 | | $B_1$ | | $K_2$ | |
| 4 | | $E_1, E_3,B_3$ | | | $A_3$ |
| 5 | | $K_1,K_3$ | | $A_1,A_2$ | |

**الشكل 36.13**: تعرّف الأشكال بشبكية مستطيلة

**مثال 8:**

استعمال شبكة خريطة الملامح الذاتية التنظيم في معطيات شجرة العبور، يمكن استعمال عينة من المعطيات المقترحة من قبل **Kohanen** عام 1989[227] لتوضيح سلوك خريطة الملامح الذاتية التنظيم. يمكن توضيح العلاقات بين النماذج تخطيطياً كما هو مبين في الشكل (38.13). تختلف النماذج المعروضة في سطر أو عمود بعضها عن بعض بمركبة واحدة فقط (بت واحد). وكذلك، توافق المسافة بين النماذج في نفس السطر أو العمود على مخطط الشكل(38.13) مباشرة المسافة الإقليدية بين الشعاعين. مثلاً، تختلف النماذج X وY المتجاورة فقط في المركبة الرابعة والمسافة الإقليدية بين (3, 3, 6, 2, 0) و(3, 3, 6, 3, 0) تساوي الواحد. المعطيات الأصلية معطاة في الشكل (37.13). بسبب هذه البنية الأنيقة، سنشير لمعلومات الاختبار هذه بشجرة العبور (**Spanning tree Data**).

قدم 32 شعاعاً (ببعد 5) الموضحة في الشكل (37.13)، بترتيب عشوائي إلى شبكة خريطة الملامح الذاتية التنظيم بطبولوجية مستطيلة الشكل مفروضة على وحداتها. كان هناك 70 وحدة صفت في مصفوفة ثنائية البعد 10 × 7 وحدة. أعطيت أسماء للنماذج لسهولة تعريف

النتائج.

| النموذج | | | | | المركبات |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | A |
| 2 | 0 | 0 | 0 | 0 | B |
| 3 | 0 | 0 | 0 | 0 | C |
| 4 | 0 | 0 | 0 | 0 | D |
| 5 | 0 | 0 | 0 | 0 | E |
| 3 | 1 | 0 | 0 | 0 | F |
| 3 | 2 | 0 | 0 | 0 | G |
| 3 | 3 | 0 | 0 | 0 | H |
| 3 | 4 | 0 | 0 | 0 | I |
| 4 | 5 | 0 | 0 | 0 | J |
| 3 | 3 | 1 | 0 | 0 | K |
| 3 | 3 | 2 | 0 | 0 | L |
| 3 | 3 | 3 | 0 | 0 | M |
| 3 | 3 | 4 | 0 | 0 | N |
| 3 | 3 | 5 | 0 | 0 | O |
| 3 | 3 | 6 | 0 | 0 | P |
| 3 | 3 | 7 | 0 | 0 | Q |
| 3 | 3 | 8 | 0 | 0 | R |
| 3 | 3 | 3 | 1 | 0 | S |
| 3 | 3 | 3 | 2 | 0 | T |
| 3 | 3 | 3 | 3 | 0 | U |
| 3 | 3 | 3 | 4 | 0 | V |
| 3 | 3 | 6 | 1 | 0 | W |
| 3 | 3 | 6 | 2 | 0 | X |
| 3 | 3 | 6 | 3 | 0 | Y |
| 3 | 3 | 6 | 4 | 0 | Z |
| 3 | 3 | 6 | 2 | 1 | 1 |
| 3 | 3 | 6 | 2 | 2 | 2 |
| 3 | 3 | 6 | 2 | 3 | 3 |
| 3 | 3 | 6 | 2 | 4 | 4 |
| 3 | 3 | 6 | 2 | 5 | 5 |
| 3 | 3 | 6 | 2 | 6 | 6 |

**الشكل 37.13**: معطيات اختبار شجرة العبور[227]

| A | B | C | D | E | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | F | | | | | | | | | | | | |
| | | G | | | | | | | | | | | | |
| | | H | K | L | M | N | O | P | Q | R | | | | |
| | | I | | | S | | | W | | | | | | |
| | | J | | | T | | | X | 1 | 2 | 3 | 4 | 5 | 6 |
| | | | | | U | | | Y | | | | | | |
| | | | | | V | | | Z | | | | | | |

**الشكل 38.13**: بنية معطيات اختبار شجرة العبور

استعملت الشبكة بأوزان أولية عشوائية. في هذا المثال، كان نصف القطر الأولي R=3، وخفض بمقدار واحد بعد كل مجموعةمن 75 تكرار. وخفض معدل التعليم خطياً من 0.6 إلى 0.01. إذا كانت الوحدة $X_{I,J}$ هي الرابحة، فإن جاراتها من الوحدات $X_{i,j}$ لكل i وj بحيث $I-R \leq i \leq I+R$ و $J-R \leq j \leq J+R$ أيضاً ستتعلم (مالم تقع قيمة i وj خارج المجال المسموح به في الطبولوجية وعدد الوحدات المختارة).

لاحظ أنه عندما تكون R = 3، فستتعلم حتـــى 49 وحدة (انظر الشكل (31.13) وعندما يكون R = 0 فإن الوحدة الرابحة فقط ستتعلم.

تبين الأشكال من (39.13) حتـــى (42.13) تطور الحل، عندما تخفض قيمة R، وفي حالة المعطيات في الشكل (37.13)، وباستعمال مصفوفة مربعة من الوحدات.

تشير بنية المعطيات المبينة في الشكل (42.13) إلى كيفية عكس توضع النماذج على وحدات التجمع علاقات شجرة العبور الأصلية فيما بين النماذج.

يمكن أن تستعمل الشبكية السداسية أيضاً في الطبولوجية الثنائية البعد. النتائج التـــي حصلنا عليها في هذه الطبولوجية مبينة في الشكل (43.13). كما في الشكل (42.13)، بنية المعطيات أيضاً، تشير إلى طريقة تأثير مكان النماذج في وحدات التجمع على شجرة العبور الأصلية. استعملت نفس خطة التكرار السابقة، 75 تكراراً لكـــل نصف قطـــر، البداية عند R = 3 والتخفيض بمقدار واحد حتـــى R = 0 .

يمكـــن مقارنـــة النتائج وشكل الخريطة وأوجه التشابه بيـــن الشكل (42.13) والشكل

(43.13) مع الشكل (38.13).

| I,J | H,K | - | G | F | - | A,B,C,D,E |
|---|---|---|---|---|---|---|
| - | - | - | - | - | - | - |
| - | - | - | - | - | - | - |
| L,M | - | - | - | - | - | - |
| S | T,U | - | - | V | - | Y,Z |
| N | - | - | - | - | X | - |
| O | - | - | - | - | - | - |
| - | - | - | 1 | - | 2 | - |
| P | - | W | - | - | - | 3 |
| Q,R | - | - | - | - | - | 4,5,6 |

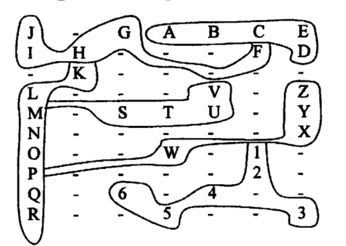
**الشكل39.13**: النتائج بعد 75 تكراراً مع R = 3

| H,I,J | - | G | A,B | F | C | D,E |
|---|---|---|---|---|---|---|
| K | - | - | - | - | - | - |
| - | - | - | - | - | - | - |
| L,M | S | - | T,U | - | - | Y,Z |
| - | - | - | V | - | - | X |
| N | - | - | - | - | - | - |
| O | - | - | 1 | - | - | - |
| P | W | - | 2 | 3 | - | - |
| - | - | - | - | - | - | - |
| Q,R | - | - | - | - | 4 | 5,6 |

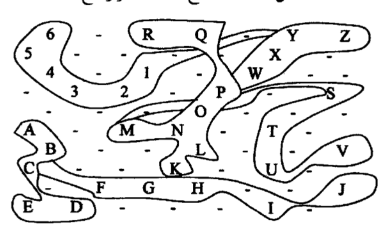
**الشكل 40.13**: النتائج بعد 75 تكراراً مع R = 2

| I,J | - | G | A | B | C | D,E |
|---|---|---|---|---|---|---|
| - | H | - | - | - | F | - |
| - | K | - | - | - | - | - |
| L | - | - | - | V | - | Z |
| M | - | S | T | U | - | Y |
| N | - | - | - | - | - | X |
| O | - | - | W | - | 1 | - |
| P | - | - | - | - | 2 | - |
| Q | - | 6 | - | 4 | - | H |
| R | - | - | 5 | - | - | 3 |

**الشكل 41.13**: النتائج بعد 75 تكراراً مع R=1




**الشكل 42.13**: النتائج بعد 75 تكراراً مع R= 0




**الشكل 43.13**: النتائج باستعمال المصفوفة السداسية

## 10.13 شبكات الانتشار المتعاكس

## Counterpropagation networks

اقترح Hecht-Nielson شبكات الانتشار المتعاكس ودرسها بين عامي 1987 و1988[48][233][234][235]، وهي شبكات متعددة الطبقات ركبت من ثلاث طبقات: طبقة الدخل وطبقة التجميع وطبقة الخرج. استعملت هذه الشبكات في مسائل ضغط المعطيات، والتطبيق العام، والتعرف، واستدعاء النماذج المترافقة.

تدرب هذه الشبكات بمرحلتين:

1. يجري في المرحلة الأولى تجميع (تكوين تجمعات) أشعة الدخل في تجمعات. في التعريف العام لشبكات الانتشار المتعاكس ليست هناك طبولوجية مفروضة للوحدات. على كل حال، إضافة طبولوجية خطية يمكن أن يحسن إنجاز الشبكة. يمكن أن تبنى المقاطع المكوَّنة إما على مسافة (متري) الجداء النقطي وإما على مسافة النظيم الإقليدي.
2. وفي المرحلة الثانية من التدريب تعدل، أو بالأحرى، تكيف الأوزان بين وحدات التجمعات ووحدات الخرج لإعطاء الاستجابة المنشودة.

سنناقش في هذا المقطع نوعين من شبكات الانتشار المتعاكس: شبكة الانتشار المتعاكس الكامل وشبكة الانتشار المتعاكس الأمامي فقط.

### 1.10.13 الانتشار المتعاكس الكامل Full counterpropagation

درست شبكة الانتشار المتعاكس الكامل وطورت لتأمين طريقة فعالة لتمثيل عدد ضخم من أزواج الأشعة (x,y). إنها تعطي التقريب (x*,y*) المبني على دخل الشعاع x (دون أية معلومات حول الشعاع y الموافق)، أو دخل الشعاع y فقط (دون أية معلومات حول x)، أو دخل الزوج (x,y) مع إمكانية حصول بعض التشويه في العناصر أو الضياع في واحد من الأشعة أو في كليهما معاً.

تَستعمل شبكة الانتشار المتعاكس الكامل أزواجَ أشعة التدريب (x,y) لتكوين تجمعات

من خلال طور التدريب الأول. في التعريف الأساسي، تختار المنافسة في طبقة التجمع (كما شرح في شبكات Kohonen) الوحدة التـــي لها دخل الشبكة الأكبر كوحدة رابحة؛ وهذا يوافق استعمال مسافة الجداء النقطي. متـــى كانت الأشعة مقارنةً باستعمال مسافة الجداء النقطي، فإنها يجب أن تجعل معيارية. ومع ذلك يمكن جعلها معيارية دون ضياع للمعلومات بإضافة مركبة إضافية، ولكن لتجنب هذا العمل ولتوفير مقارنة مباشرة بين شبكة الانتشار المتعاكس الكامل وشبكة الانتشار المتعاكس الأمامي فقط، سنستعمل النظيم الإقليدي للشبكتين (مثلما استُخدم من قبل في شبكة خريطة الملامح ذاتية التنظيم وشبكة تعليم التكميم الشعاعي).

بنية الشبكة موضحة في الشكل (44.13)، حيث تبين الأشكال (45.13) و(46.13) الوحدات الفعالة خلال كل من طوري تدريب شبكة الانتشار المتعاكس الكامل مع وصف للأوزان.

### 1.1.10.3 خوارزمية تدريب شبكة الانتشار المتعاكس الكامل

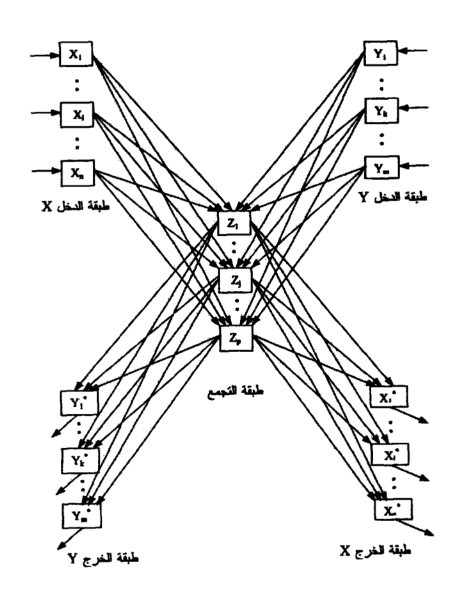
**Training full counterpropagation network**

كما ذكرنا من قبل، تتضمن خوارزمية التدريب لشبكة الانتشار المتعاكس طورين اثنين. خلال الطور الأول تكون الوحدات في طبقة الدخل X وطبقة التجمع وطبقة الدخل Y فعالة، وتتنافس الوحدات في طبقة التجمع. لم تظهر على الشكل (44.13) الوصلات الداخلية فيما بينها. في الشبكة الأساسية للانتشار المتعاكس ليست هناك طبولوجية مفروضة لوحدات طبقة التجمع، أي يسمح للوحدة الرابحة فقط بالتعلم.

تعطى قاعدة التعليم لتحديث الأوزان على وحدة طبقة التجمع الرابحة كما يلي:

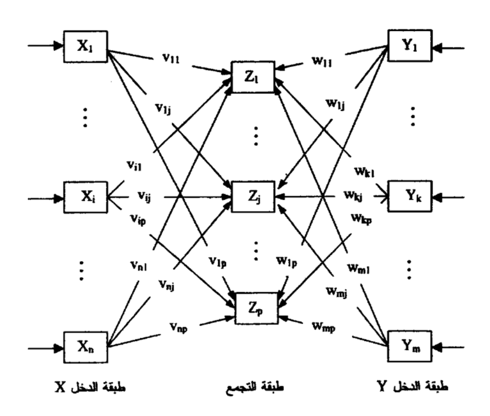
$$v_{iJ}^{new} = (1-\alpha)v_{iJ}^{old} + \alpha x_i \quad i = 1,2,\cdots,n$$
$$w_{kJ}^{new} = (1-\beta)w_{kJ}^{old} + \beta y_k \quad k = 1,2,\cdots,m \qquad (46.13)$$

وهذا هو تعليم Kohonen الأساسي، الذي يتضمن التنافس بين الوحدات وتحديث أوزان الوحدة الرابحة.




**الشكل 44.13:** بنية شبكة الانتشار المتعاكس الكامل

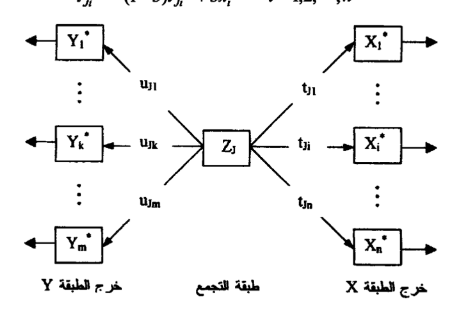
خلال الطور الثانـــي من الخوارزمية، تبقى الوحدة J فقط فعالة في طبقة التجمع. وتكون الأوزان من وحدة التجمع الرابحة J إلى وحدات الخرج بنوعيه X وY معدلة بحيث يكون شعاع تفعيلات الوحدات في طبقة الخرج **y*** قريـــباً من شعاع الدخل **y**، ويكون **x*** قريباً من **x**.




**الشكل 45.13:** الوحدات الفعالة خلال طور تدريب الانتشار المتعاكس

يعطى تحديث أوزان الوحدات في طبقات الخرج Y وX كما يلي:

$$u_{Jk}^{new} = (1-a)u_{Jk}^{old} + a\,y_k \quad k = 1,2,\cdots,m$$
$$t_{Ji}^{new} = (1-b)t_{Ji}^{old} + bx_i \quad i = 1,2,\cdots,n \qquad (47.13)$$




**الشكل 46.13:** الطور الثانـــي من تدريب الانتشار المتعاكس

وهذا معروف بتعليم Grossberg الذي استعمل هنا كحالة خاصة من تعليم outstar العام أكثر (Hecht-Neilsen عام 1990[46]).

يحدث تعليم outstar لكل الوحدات في طبقة خاصة، وهذا يعنـــي أنه ليست هناك منافسة فيما بين هذه الوحدات. على أية حال، تكون أشكال تحديث الأوزان بتعليم Kohonen وتعليم Grossberg قريبة جداً بعضها من بعض.

يمكن أن ينظر إلى قواعد التعليم لطبقات الخرج كتعليم قاعدة دلتا. وللتأكد من ذلك، نفترض $y_k$ هو قيمة الخرج المنشود للوحدة $Y_k^*$ و $u_{Jk}^{old}$ هو التفعيل المحسوب للوحدة (بافتراض أن الإشارة المرسلة بواسطة الوحدة $Z_J$ تساوي الواحد). سيكون لدينا الآن:

$$\begin{aligned} u_{Jk}^{new} &= (1-a)u_{Jk}^{old} + ay_k \\ &= u_{Jk}^{old} + a(y_k - u_{Jk}^{old}) \end{aligned} \qquad (48.13)$$

وهكذا، تغير الوزن هو ببساطة معدل التعليم a مضروباً في الخطأ. تماماً نفس الملاحظات مطبقة على تحديث أوزان الوحدات في طبقة الخرج X ( Dayhoff عام 1990[188]).

قبل مناقشة إجراءات خوارزمية تعليم الانتشار المتعاكس سنقوم بتلخيص بعض التعاريف:

**x**: شعاع دخل التدريب

$$\mathbf{x} = x_1, x_2, \ldots, x_n$$

**y**: الخرج المنشود الموافق للدخل **x**

$$\mathbf{y} = (y_1, y_2, \cdots, y_n)$$

$\mathbf{z_j}$: تفعيل وحدة طبقة التجمع

**x***: التقريب المحسوب للشعاع **x**

**y***: التقريب المحسوب للشعاع **y**

$v_{ij}$ : الوزن من طبقة الدخل X، الوحدة $X_i$، إلى طبقة التجمع، الوحدة $Z_j$.

$w_{kj}$ : الوزن من طبقة الدخل Y، الوحدة $Y_k$، إلى طبقة التجمع، الوحدة $Z_j$.

$u_{jk}$ : الوزن من طبقة التجمع، الوحدة $Z_j$، إلى طبقة الخرج Y، الوحدة $Y_k^*$.

$t_{ji}$ : الوزن من طبقة التجمع، الوحدة $Z_j$، إلى طبقة الخرج X، الوحدة $X_i^*$.

$\alpha$ و$\beta$: معدلات التعليم للأوزان الداخلة إلى طبقة التجمع (تعليم Kohonen).

a وb : معدلات التعليم للأوزان الصادرة عن طبقة التجمع (تعليم Grossberg).

يمكن عرض خوارزمية تعليم الانتشار المتعاكس الكامل كما يلي:

1. إعطاء الأوزان ومعدلات التعليم قيماً أولية،. .الخ.
2. مادام شرط التوقف للطور الأول غير محقق كرر الخطوات من 3-8
3. لكل زوج دخل تدريب (x,y)، كرر الخطوات من 4-6
4. ضع تفعيلات طبقة الدخل X مساوية للشعاع x، ضع تفعيلات طبقة الدخل Y مساوية للشعاع y
5. أوجد وحدة التجمع الرابحة: معرفة دليلها J
6. حدث أوزان الوحدة $Z_J$:

$$v_{iJ}^{new} = (1-\alpha)v_{iJ}^{old} + \alpha x_i \quad i = 1,2,\cdots,n$$
$$w_{kJ}^{new} = (1-\beta)w_{kJ}^{old} + \beta y_k \quad k = 1,2,\cdots,m$$

7. خفض قيمة معدلات التعليم $\alpha$ و$\beta$.
8. اختيار شرط التوقف لطور التدريب الأول.
9. مادام شرط توقف الطور الثانـــي للتدريب غير محقق، كرر الخطوات من 10-16 ($\alpha$ و$\beta$ لهما قيم صغيرة ثابتة خلال الطور الثاني)
10. لكل زوج (x,y) دخل تدريب، كرر الخطوات من 11-14
11. ضع تفعيلات طبقة الدخل X مساوية للشعاع x، ضع تفعيلات طبقة الدخل Y مساوية للشعاع y
12. أوجد وحدة التجمع الرابحة: معرفة دليلها J
13. حدث أوزان الوحدة $Z_J$:

$$v_{iJ}^{new} = (1-\alpha)v_{iJ}^{old} + \alpha x_i \quad i = 1,2,\cdots,n$$
$$w_{kJ}^{new} = (1-\beta)w_{kJ}^{old} + \beta y_k \quad k = 1,2,\cdots,m$$

14. تحديث الأوزان من الوحدة $Z_J$ إلى طبقات الخرج:

$$u_{Jk}^{new} = (1-a)u_{Jk}^{old} + a\, y_k \quad k = 1,2,\cdots,m$$
$$t_{Ji}^{new} = (1-b)t_{Ji}^{old} + b x_i \quad i = 1,2,\cdots,n$$

15. خفض معدلات التعليم a و b

16. اختبر شرط توقف الطور الثاني

في الخطوتين 5 و12:

خذ الوحدة ذات الدليل الأخفض. لاستعمال مسافة الجداء النقطي، أوجد وحدة التجمع Zj ذات دخل الشبكة الأكبر:

$$z(net_j) = \sum_i x_i v_{ij} + \sum_k y_k w_{kj} \qquad (49.13)$$

ستكون أشعة الوزن وأشعة الدخل معيارية لكي تستعمل في مسافة الجداء النقطي. لاستعمال المسافة الإقليدية، أوجد وحدة التجمع Zj، التـــي يكون مربع مسافتها عن أشعة الدخل أصغرياً، وتعطى بالعلاقة:

$$D_j = \sum_i (x_i - v_{ij})^2 + \sum_k (y_k - w_{kj})^2 \qquad (50.13)$$

بعد أن تُدرَّب الشبكة العصبونية الصنعية ذات الانتشار المتعاكس الكامل، يمكن استعمالها لإيجاد التقريبات **x*** و**y*** لزوج شعاعي الدخل **x** و**y**.

أشار Hecht-Nielson عام 1991[115] إلى هذه العملية بالنمو (accretion)، بأسلوب معاكس للاستيفاء الداخلي (التوليد) بين القيم المعروفة للتابع.

ستكون إجراءات التطبيق على النحو التالي:

1. وضع القيم الأولية للأوزان.
2. لكل زوج تدريب (**x,y**)، كرر الخطوات من 3 إلى 5.
3. ضع تفعيلات طبقة الدخل X مساوية للشعاع **x**، ضع تفعيلات طبقة الدخل Y مساوية للشعاع **y**
4. أوجد وحدة التجمع ZJ التـــي تكون أقرب لزوج الدخل
5. حساب التقريب إلى **x** و**y**:

$$\begin{aligned} x_i^* &= t_{Ji} \\ y_k^* &= u_{Jk} \end{aligned} \qquad (51.13)$$

يمكن أن تستعمل الشبكة أيضاً في حالة الاستيفاء الداخلي. في هذه الحالة، يسمح لعدة وحدات أن تكون فعالة في طبقة التجمع. توضع التفعيلات بحيث يتحقق $\sum_j z_j = 1$ (لتشكيل

تركيب محدب للقيم). يعطى تقريب الاستيفاء الداخلي لـــ x و y بما يلي:

$$x_i^* = \sum_j z_j t_{ji}$$
$$y_k^* = \sum_j z_j u_{jk} \qquad (52.13)$$

تزداد دقة التقريب باستعمال الاستيفاء الداخلي. في حالة الاختبار بشعاع دخل واحد فقط x (هذا يعنـــي، عدم وجود معلومات حول y)، من المفضل إيجاد الوحدة الرابحة J المبنية على مقارنة الشعاع x فقط وأول n مركبة من شعاع الوزن لكل وحدة طبقة تجمع.

**مناقشة:**

تكون مصفوفة الوزن V من طبقة الدخل X إلى طبقة التجمع على الأغلب مماثلة لمصفوفة الوزن T من طبقة التجمع إلى طبقة الخرج X. وبالمثل، تكـــون مصفوفات الوزن W و U لـــ Y و *Y أيضاً وبوجه أساسي متماثلة. ولكن هذه الحالات من المتوقع تحققها، لأن شكل قواعد التعليم ستكون نفسها ونفس معدلات التعليم الأولية المستعملة.

توضِّح الفروقات الضئيلة في هذه المصفوفات حقيقة أن بعض النماذج يمكن أن تكون متعلمة من قبل وحدة واحدة في بدايات التدريب (لطبقة التجمع)، لكن في نهاية المطاف يمكن أن تكون متعلمة مـــن قبل وحـــدة مختلفة. هذه "الهجرة" لا تؤثر في تعليم طبقة الخرج (المصفوفة T والمصفوفة U).

هناك عامل آخر في إحداث الفروق في مصفوفات الأوزان هو التعليم الإضافي (عند معدل منخفض جداً) الذي يحدث للمصفوفات V و W خلال تكوين المصفوفات U و T.

**مثال 9:**

شبكة الانتشار المتعاكس الكامل للتابع $y = 1/x$، في هذا المثال سنحاول اختبار إنجاز شبكة الانتشار المتعاكـــس بغيـــة حساب قيمة التابع $y = 1/x$ على المجال [10–0.1]. لنفترض أن لدينا 10 وحدات تجمع (في طبقة Kohonen)؛ وهناك وحدة واحدة في طبقة الدخل X، ووحدة واحدة في طبقة الدخل Y، ووحدة واحدة في طبقة الخرج X، ووحدة واحدة في طبقة الخرج Y.

لنفترض كذلك أن لدينا عدداً كبيراً من نقاط التدريب (1000 مثلاً)، لقيم x بين

[0.1–10] وقيم y الموافقة تعطى بواسطة التابع $y = 1/x$. إن نقاط دخل التدريب، التـــي تكون موزعة توزيعاً منتظماً على طول المنحني، ستقدم بترتيب عشوائي.

إذا أحسنا اختيار القيم الأولية للأوزان (على وحدات التجمع)، عندئذ بعد أول طور تدريب، ستتوزع وحدات التجمع توزيعاً منتظماً على طول المنحنـــي. وإذا استعملنا بنية طبولوجية خطية (كما هي الحالة في شبكة خريطة الملامح ذاتية التنظيم) على وحدات التجمع، فإن هذا سيحسن فرص تمثيل الأوزان للنقاط على المنحنـــي في أسلوب أمثلي إحصائي.

تعطي النتائج النموذجية القيم التالية لأوزان وحدات التجمع. بالطبع، يمكن أن يفسر هذا كأمكنة في المستوي (x,y) التـــي تمثل وحدات التجمع. الوزن الأول لكل وحدة تجمع هو الوزن من وحدة الدخل X، والثانـــي هو من وحدة الدخلY، الأوزان هي :

| وحدة التجمع | v | W |
|---|---|---|
| $Z_1$ | 0.11 | 9.0 |
| $Z_2$ | 0.14 | 7.0 |
| $Z_3$ | 0.20 | 5.0 |
| $Z_4$ | 0.30 | 3.3 |
| $Z_5$ | 0.6 | 1.6 |
| $Z_6$ | 1.6 | 0.60 |
| $Z_7$ | 3.3 | 0.0 |
| $Z_8$ | 5.0 | 0.20 |
| $Z_9$ | 7.0 | 0.14 |
| $Z_{10}$ | 9.0 | 0.11 |

بعد طور التدريب الثاني، ستكون أوزان وحدات الخرج تقريباً نفس الأوزان المعطاة إلى وحدات التجمع. الأوزان موضحة على مخطط الشبكة الكلي (47.13) وقيمة التابع على الشكل (48.13).

نستطيع استعمال هذه الشبكة للحصول على قيمة تقريبية لـــ y في حالة $x = 0.12$ كما يلي:

1. وضع الأوزان بقيم أولية

2. في حالة الدخل $x = 0.12$ و$y = 0.0$، كرر الخطوات من 3-5

3. ضع تفعيلات طبقة الدخل X مساوية للشعاع **x**، ضع تفعيلات طبقة الدخل Y مساوية للشعاع **y**

4. إيجاد الدليل J لوحدة التجمع الرابحة: تعطى مربعات المسافات من الوحدة إلى كل من وحدات التجمع كما يلي:

$$D_1 = (0.12 - 0.11)^2 + (0.00 - 9.00)^2 = 81$$
$$D_2 = (0.12 - 0.14)^2 + (0.00 - 7.00)^2 = 49$$
$$D_3 = (0.12 - 0.20)^2 + (0.00 - 5.00)^2 = 25$$
$$D_4 = (0.12 - 0.30)^2 + (0.00 - 3.30)^2 = 11$$
$$D_5 = (0.12 - 0.60)^2 + (0.00 - 1.60)^2 = 2.8$$
$$D_6 = (0.12 - 1.60)^2 + (0.00 - 0.60)^2 = \underline{2.6}$$
$$D_7 = (0.12 - 3.30)^2 + (0.00 - 0.30)^2 = 10.2$$
$$D_8 = (0.12 - 5.0\ 0)^2 + (0.00 - 0.20)^2 = 23.9$$
$$D_9 = (0.12 - 7.00)^2 + (0.00 - 0.14)^2 = 47.4$$
$$D_{10} = (0.12 - 9.00)^2 + (0.00 - 0.11)^2 = 78.9$$

المحسوبة على الدخل الكلي، وهكذا فإن وحدة التجمع الأقرب هي $J = 6$.

5. احسب تقريب **x** و**y**:

$$x^* = t_J = 1.6$$
$$y^* = u_J = 0.6$$

من الواضح أن، هذا التقريب ليس التقريب الذي نرغب بإيجاده، وبسبب أن المعلومات المتوفرة لدينا تتعلق فقط بالدخل **x**، سنستعمل التعديل المذكور من قبل لإجراء المسألة. وهكذا، إذا بنينا بحثنا على وحدة التجمع الرابحة التـــي تقع على مسافة من الدخل **x** إلى الوزن الموافق لكل وحدة تجمع، سنجد القيم التالية في الخطوتين 4 و5:

4. إيجاد الدليل J لوحدة التجمع الرابحة: تعطى مربعات المسافات من الوحدة إلى كل من وحدات التجمع كما يلي:

$D_1 = (0.12 - 0.11)^2 = \underline{\mathbf{0.0001}}$

$D_2 = (0.12 - 0.14)^2 = 0.0004$

$D_3 = (0.12 - 0.20)^2 = 0.064$

$D_4 = (0.12 - 0.30)^2 = 0.032$

$D_5 = (0.12 - 0.60)^2 = 0.23$

$D_6 = (0.12 - 1.60)^2 = 2.2$

$D_7 = (0.12-3.30)^2 = 10.1$

$D_8 = (0.12-5.0\ 0)^2 = 23.8$

$D_9 = (0.12-7.00)^2 = 47.3$

$D_{10} = (0.12-9.00)^2 = 81$

المحسوبة على الدخل من x فقط، وحدة التجمع الأقرب هي J=1

5.احسب تقريب **x** و**y**:

$\mathbf{x}^* = t_J = 0.11$

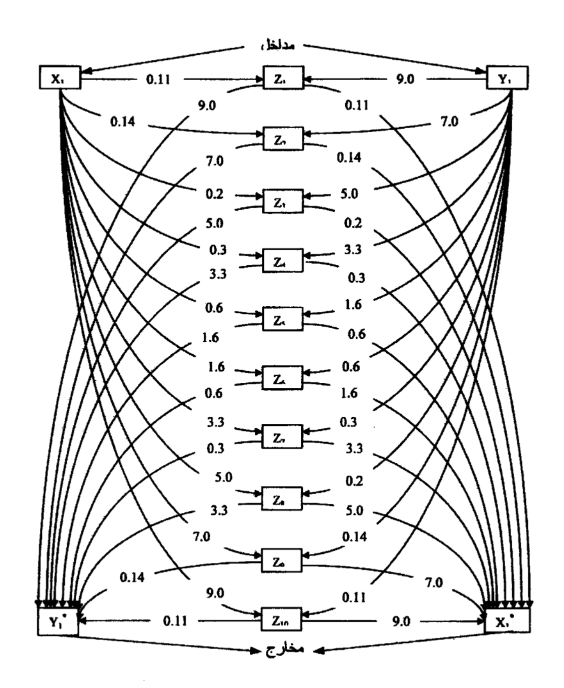
$\mathbf{y}^* = u_J = 9.00$

## 2.10.13 شبكة الانتشار المتعاكس الأمامي فقط

**Forward-only counterpropagation**

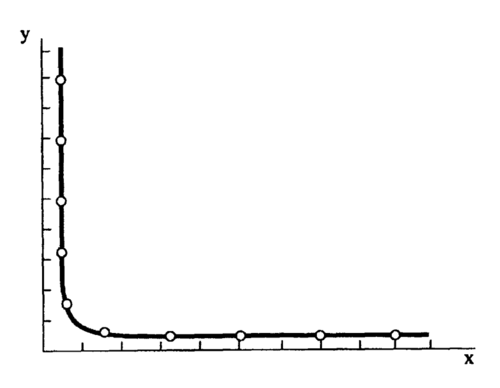
إن شبكة الانتشار المتعاكس الأمامي فقط هي نسخة مبسطة عن شبكة الانتشار المتعاكس الكامل. صممت هذه الشبكة المبسطة لتقريب التابع **y = f(x)** الذي لا يفترض فيه أن يكون عكوساً؛ أي إن شبكات الانتشار المتعاكس الأمامي هي الوحيدة التـــي يمكن أن تستعمل إذا كان التطبيق من **x** إلى **y** معرفاً فقط (لذا سميت بشبكة الانتشار المتعاكس الأمامي فقط) وليس التطبيق من **y** إلى **x** (كما رأينا في شبكة الانتشار المتعاكس الكامل، حيث يكون التطبيق من **x** إلى **y** عكوساً، ومن هنا اشتق اسم هذه الشبكات بالانتشار المتعاكس الكامل).

تختلف هذه الشبكة عن باقي شبكات الانتشار المتعاكس الكامل باستعمال الأشعة **x** فقط لتشكيل تجمعات على وحدات **Kohonen** خلال الطور الأول من التدريب.




**الشكل 47.13**: شبكة الانتشار المتعاكس الكاملة للتابع $y = 1/x$

يَستعمل التمثيل الأساسي للانتشار المتعاكس الأمامي فقط المسافةَ الإقليدية بين شعاع الدخل وشعاع الوزن (الأنموذج) في وحدة Kohonen (عوضاً عن مسافة الجداء النقطي المستعملة في الانتشار المتعاكس الكامل الأساسي). على أية حال، يمكن استعمال أي مسافة في أي شبكة انتشار متعاكس.




**الشكل 48.13:** منحنـــي التابع $y = 1/x1$ مبيناً مكان وضع وحدات التجمع العشر.

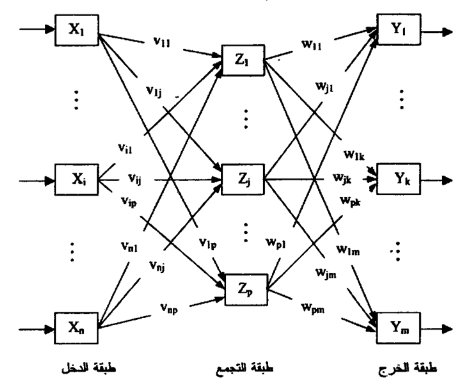
بنية شبكة الانتشار المتعاكس الأمامي موضحة في الشكل (49.13)، حيث تظهر مشابهة تماماً لشبكة الانتشار المتعاكس الكامل لأن الوصلات الجانبية فيما بين وحدات طبقة التجمع لم تظهر في الشكل. بوجه عام، في الانتشار الأمامي فقط، وبعد المنافسة، ستكون وحدة واحدة فقط في طبقة التجمع فعالة وسترسل إشارتها إلى طبقة الخرج.

### 1.2.10.13 خوارزمية تعليم شبكة الانتشار المتعاكس الأمامي

تمر إجراءات تدريب هذه الشبكة بمراحل عديدة. أولاً، يقدم شعاع الدخل إلى وحدات الدخل. ثم تتنافس وحدات طبقة التجمع (في منافسة الرابح يحوز الكل) لتفوز واحدة بتعلم شعاع الدخل. بعد أن تقدم كل أشعة التدريب إلى الشبكة، تخفض قيمة معدل التعليم وتقدم الأشعة ثانية؛ يستمر تكرار هذا الإجراء مرات عديدة. بعد أن تكون الأوزان من طبقة الدخل إلى طبقة التجمع قد دربت (معدل التعليم خفض إلى قيمة صغيرة جداً)، تدرَّب الأوزان من طبقة التجمع إلى طبقة الخرج.

الآن، عند تقديم كل شعاع دخل تدريب إلى طبقة الدخل، يقدم شعاع الهدف المنشود المرافق إلى طبقة الخرج. ترسل وحدة التجمع الرابحة، ولتكن J، إشارةَ 1 إلى طبقة الخرج.

لكل وحدة خرج k إشارة دخل محسوبة $w_{Jk}$ وإشارة هدف منشود $y_k$. باستعمال الفرق بين هاتين القيمتين، يجري تحديث الأوزان بين وحدة التجمع الرابحة وطبقة الخرج. إن قاعدة تعليم تحديث الأوزان مشابهة لقاعدة تعليم الأوزان من وحدات الدخل إلى وحدات التجمع.




**الشكل 49.13**: شبكة الانتشار المتعاكس الأمامي فقط

**قاعدة التعليم للأوزان من وحدات الدخل إلى وحدات التجمع:**

$$v_{iJ}^{new} = v_{iJ} + \alpha(x_i - v_{iJ})$$
$$v_{iJ}^{new} = (1-\alpha)v_{iJ}^{old} + \alpha\, y_k \qquad (53.13)$$

**قاعدة التعليم للأوزان من وحدات التجمع إلى وحدات الخرج:**

$$w_{Jk}^{new} = w_{Jk} + a(y_k - w_{Jk})$$
$$w_{Jk}^{new} = (1-a)w_{Jk}^{old} + a\, x_i \qquad (54.13)$$

على أية حال، إذا فسر $w_{Jk}$ كخرج محسوب (أي، $y_k = w_{Jk}$)، وكان تفعيل وحدات التجمع مشمولاً، أي:

$z_j = 1$ في حالة $j = J$ و$z_j = 0$ ما عدا ذلك.

عندئذ يمكن أن تكتب قاعدة التعليم للأوزان من وحدات التجمع إلى وحدات الخرج

على شكل قاعدة دلتا:

$$w_{jk}^{new} = w_{jk} + az_j(y_k - w_{Jk}) \qquad (55.13)$$

يستمر تدريب الأوزان من وحدات الدخل إلى وحدات التجمع عند معدل تعليم منخفض، على حين يجري تخفيض معدل تعليم الأوزان من وحدات التجمع إلى وحدات الخرج تدريجياً.

قبل استعراض الخوارزمية سنعرف بعض المصطلحات:

$\alpha$ وa: وسطاء معدل التعليم، $0.5 < \alpha < 0.8$ و$0 < a < 1$.

اقترح Hecht-Neilsen عام 1988[48] القيم التالية: $\alpha = 0.6$ و$a = 1$.

**X**: شعاع تفعيل وحدات الدخل، حيث $\mathbf{x} = (x_1, x_2, \ldots, x_n)$

$\|\mathbf{x} - \mathbf{v}\|$: المسافة الإقليدية بين الشعاعين **x** و**v**.

كما في شبكات الانتشار المتعاكس الكامل، ليس هناك بنية طبولوجية مفترضة لوحدات التجمع في الصيغ الأساسية لشبكات الانتشار المتعاكس. في حالات عديدة، يمكن أن يتحسن التدريب باستعمال بنية خطية على وحدات التجمع. تساعد البنية على التأكد، بعد التدريب، من كون أوزان وحدات التجمع موزعة بأسلوب أمثلي إحصائياً.

خوارزمية تدريب شبكة الانتشار المتعاكس الأمامية فقط هي التالية:

1. إعطاء الأوزان ومعدلات التعليم قيماً أولية،. .الخ.
2. مادام شرط التوقف للطور الأول غير محقق كرر الخطوات من 3-8
3. لكل دخل تدريب **x**، كرر الخطوات من 4-6
4. ضع تفعيلات طبقة الدخل X مساوية للشعاع **x**
5. أوجد وحدة التجمع الرابحة: معرفة دليلها J
6. حدث أوزان الوحدة الرابحة $Z_J$:

$$v_{iJ}^{new} = (1 - \alpha)v_{iJ}^{old} + \alpha x_i \quad i = 1,2,\cdots,n$$

7. خفض قيمة معدلات التعليم $\alpha$
8. اختبار شرط التوقف لطور التدريب الأول.
9. مادام شرط توقف الطور الثانـــي للتدريب غير محقق، كرر الخطوات من 10-16 ( $\alpha$ له

قيم صغيرة ثابتة خلال الطور الثاني).

10. لكل زوج (**x,y**) دخل تدريب، كرر الخطوات من 11-14.

11. ضع تفعيلات طبقة الدخل X مساوية للشعاع **x**، ضع تفعيلات طبقة الدخل Y مساوية للشعاع **y**.

12. أوجد وحدة التجمع الرابحة: معرفة دليلها J

13. حدث الأوزان إلى الوحدة $Z_J$ ( $\alpha$ صغير):

$$v_{iJ}^{new} = (1-\alpha)v_{iJ}^{old} + \alpha x_i \quad i = 1,2,\cdots,n$$

14. تحديث الأوزان من الوحدة $Z_J$ إلى طبقات الخرج:

$$w_{Jk}^{new} = (1-a)w_{Jk}^{old} + a\, y_k \quad k = 1,2,\cdots,m$$

15. خفض معدلات التعليم a وb

16. اختبر شرط توقف الطور الثاني

في الخطوتين 5 و12:

خذ الوحدة ذات الدليل الأخفض. لاستعمال مسافة الجداء النقطي، أوجد وحدة التجمع $Z_j$ ذات دخل الشبكة الأكبر:

$$z(net_j) = \sum_i x_i v_{ij} \qquad (56.13)$$

يجب أن تكون أشعة الوزن وأشعة الدخل معيارية لكي تستعمل في مسافة الجداء النقطي.

لاستعمال المسافة الإقليدية، أوجد وحدة التجمع $Z_j$ التـــي مربع مسافتها عن نموذج الدخل:

$$D_j = \sum_i (x_i - v_{ij})^2$$

يكون أصغرياً.

ستكون إجراءات التطبيق على النحو التالي:

1.ضع القيم الأولية للأوزان (بالتدريب كما ذكر فيما سبق)

2. لكل شعاع دخل تدريب **x**،كرر الخطوات من 3-5

3. ضع تفعيلات طبقة الدخل X مساوية للشعاع **x**

4. أوجد وحدة التجمع $Z_J$ التـــي تكون أقرب لشعاع الدخل **x**

5. ضع تفعيلات وحدات الخرج

$$y_k = w_{Jk}$$

يمكن أن تستعمل الشبكة أيضاً في حالة الاستيفاء الداخلي. في هذه الحالة، يسمح لعدة وحدات أن تكون فعالة في طبقة التجمع. توضع التفعيلات بحيث تحقق $\sum_j z_j = 1$ (لتشكيل تركيب محدب للقيم). يعطى تفعيل وحدات الخرج بما يلي:

$$y_k = \sum_j z_j w_{jk} \qquad (57.13)$$

وتزداد دقة التقريب باستعمال الاستيفاء الداخلي.

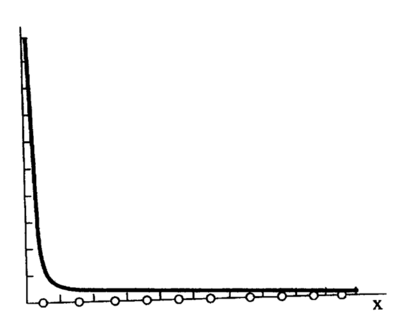
**مثال 10:**

شبكة الانتشار المتعاكس الأمامية فقط للتابع y = 1/x، في هذا المثال سنحاول اختبار إنجاز شبكة الانتشار المتعاكس الأمامية فقط بغية حساب قيمة التابع y = 1/x على المجال [0.1–10].

لنفترض أن لدينا 10 وحدات تجمع (في طبقة Kohonen)، وهناك وحدة واحدة في طبقة الدخل X، ووحدة واحدة في طبقة الخرج Y. لنفترض كذلك أن لدينا عدداً كبيراً من نقاط التدريب (1000 مثلاً)، لقيم X بين [0.1–10]. تكون نقاط دخل التدريب موزعة توزيعاً منتظماً على طول المنحنـــي، وستقدم بترتيب عشوائي.

بعد الطور الثانـــي من التدريب، ستكون الأوزان إلى وحدات الخرج Y بشكل تقريبـــي 5.5, 0.75, 0.4,....., 0.1. وهكذا، سيكون التقريب لقيم التابع في حالة قيم كبيرة لـــ x أكثر دقة في حالة قيم صغيرة.

يوضح الشكل (50.13) الأوزان المرافقة مع كل وحدة تجمع. قارن هذه النتائج مع نتائج المثال السابق (في شبكة الانتشار المتعاكس الكامل)، نلاحظ أنه حتـــى إذا صممت الشبكة بغية تقريب التطبيق من x إلى y فقط، فإن شبكة الانتشار المتعاكس الكامل توزع وحدات التجمع بأسلوب يعطي تقريباً أكثر دقة عبر المجال الكامل لقيم الدخل.




**الشكل 50.13**: نتائج التابع $y = 1/x$ باستعمال شبكة الانتشار المتعاكس الأمامية فقط

## 11.13 تطبيقات شبكات التكميم الشعاعي وخريطة الملامح الذاتية التنظيم

## Application of VQ and SOFM

كما ذكرنا من قبل، إن لشبكات التكميم الشعاعي تطبيقات واسعة في مجال تكميم الشعاع لضغط المعطيات، وتصحيح الخطأ، وتوليد كلمات الرموز(الترميز). وباعتبار أن شبكات خريطة الملامح الذاتية التنظيم هي تعميم لشبكات التكميم الشعاعي، لذا يمكن استخدامها في هذه التطبيقات. بالإضافة إلى أن شبكات خريطة الملامح ذاتية التنظيم استعملت بفعالية في مسائل الاستمثال (كحل مسألة البائع الجوال وسواها)، وفي التحكم، وتعرف الكتابة اليدوية، وتمييز الأشكال، وتعرف إشارة الكلام، ....الخ. في هذه الفقرة سنصف تطبيقين لنموذجين واعدين جداً من بين هذه التطبيقات.

### 1.11.13 الآلة الكاتبة اللفظية The phonetic typewriter

طور Teuvo Kohonen وزملاؤه في جامعة Helsinki للتكنولوجيا، آلة كاتبة لفظية منذ أوائل عام 1980. هذه الآلة الكاتبة اللفظية عبارة عن نظام تمييز يستطيع تدوين كلام غير محدد إلى نص صحيح مسقط عمودياً. عندما ينجز هذا المشروع نهائياً للغات متعددة، وينفذ

صناعياً سيكون له تأثير كبير في معالجة المعلومات. سيثير النص المؤتمت من لفظ إلى نص مكتوب ضجة كبيرة في الأوساط المكتبية، فضلاً عن أعمال النشر.

وقد حققت مجموعة Kohonen في جامعة Helsinki درجة عالية من النجاح في اللغات الصوتية كاليابانية والفلندية، ولكن ليس من الواضح بعد كيف يمكن توسيع نتائجهم بسهولة لتشمل اللغات الأخرى، كاللغات الأقل صوتية مثل الإنكليزية والروسية أو حتـــى الصينية، حيث ينبغي مراعاة النغمة.

نفّذت تجارب باستعمال ثلاثة ذكور فيلنديين متكلمين عاديين بدقة إنجاز وصلت إلى ما بين 91-96%، حيث نفذت التجارب بأربع مرات تكرار لمجموعة التدريب المؤلفة من 311 كلمة. احتوت كل مجموعة من الكلمات 1737 مقطع صوتـــي (phonemes).

استُعملت شبكة LVQ تقليدية لتطبيق المقاطع الصوتية إلى صفوف مرجعية (سجل الرموز). دخل شبكة LVQ هو مركبات طيفية (cepstral) محسوبة من طيف الكلام (لاحَظ Bogert & Healy & Tukey عام 1963[236] أن لوغارتم طيف القدرة لإشارة ما يحتوي على صدى (echo) له مركبات دورية إضافية تبعاً للصدى، ومن ثَم فإن تحويل فورييه للوغارتم طيف القدرة سيظهر قيمة عظمى عند صدى متأخر، وأُطلق على ذلك اسم cepstrum. اشتق أصل الكلمة من كلمة الطيف spectrum ولكنهم بدلوا بين أحرف هذه الكلمة لأننا، بوجه عام، نجد أنفسنا نعمل في المجال الترددي بالطرائق المعتاد العمل بها في المجال الزمنـــي والعكس صحيح). مجموعة المعاملات الطيفية هي تركيب غير خطي لتحويلين متتابعين لفورييه (اشتقت من 256 نقطة تحويل ومن ثم ضغطت). صنفت أشعة المعاملات الناتجة إلى واحد من 26 صفاً صوتياً لأحرف اللغة الفلندية.

استعمل التعليم بمعلم باستعمال LVQ1 وLVQ2 LVQ3 للتكميم وجرت مقارنة النتائج في مستوى الدقة مع طرائق القرار الإحصائية غير الخطية. حُسِب شعاع نموذج جديد من الطيف كل 10 ميلي ثانية، أنجز التصنيف لشعاع النموذج في الزمن الحقيقي باستعمال مكونات وحواسيب شخصية متوفرة تجارياً.

ينتج هذا التصنيف متتالية من الرموز أو أشباه المقاطع الصوتية، حيث تدمج في مقاطع يمثل كل منها مقطعاً صوتياً في الكلام. بالطبع هذا يتطلب تحليلاً إحصائياً تُوضع فيه خطة اقتراع.

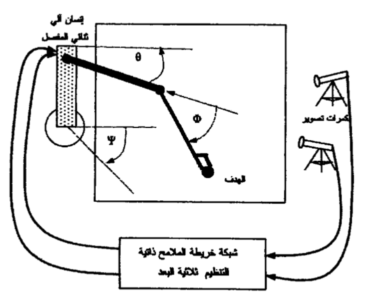
استعملت طريقة تستخدم الآلاف من القواعد تدعى DFC؛ (Dynamically Focusing Context) لإتمام دمج المقاطع الصوتية وتشكيل المقاطع.

هناك أيضاً طرق معروفة جيداً استُعملت لعملية الترميز بما في ذلك نموذج ماركوف المخفي HMM (Hidden Markov model)، ثم استُعملت طريقة DEC الرمزية؛ (Dynamically Expanding Context) لتصحيح الأخطاء الصوتية.

إن مسألة تحويل كلام إلى نص مسألة معقدة جداً؛ فقد قدر أن تعريف شخصية متكلم تعريفاً صحيحاً، بكلام مستمر باستعمال 20000 كلمة لغوية في المعجم يتطلب قدرة حساب 100000 MIPS ( مئة ألف مليون تعليمة في الثانية الواحد)، وهذا يكافئ 100 حاسوب متطور موصولة على التوازي (Reddy وZue عام 1983[28]). لذا، من غير المدهش بعد ذلك أن ندرك أن حل مثل هذه المسألة ليس أمراً سهلاً.

## 2.11.13 التحكم في الإنسان الآلي Robot control

لتوضيح مقدرة شبكة خريطة الملامح الذاتية التنظيم في إنجاز التطبيقات غير الخطية، سنناقش تطبيق ذراع إنسان آلي معقد. لنفترض أن الذراع له ثلاث زوايا حرة كما هو موضح في الشكل(51.13).




الشكل 51.13: تحكم في الإنسان الآلي

ما نريده هو تطبيق مكان الهدف المستشعر به بواسطة كمرتـــي تصوير إلى مخارج ثلاثة

محركات قيادة تقوم بوضع المقبض عند مكان الهدف المنشود.

توفر الكاميرتان، بسبب الفصل المكانـــي لهما، معلومات مصورة للشبكة. يوافق كل مكان هدف نقطة في مستوي الصورة لكل من الكاميرتين. تسجل الكامرتان الإحداثيات العمودية للهدف. عندما تؤخذ مخارج الاحداثيات الشعاعية الثنائية البعد من الكاميرتين، فإنها توفر معلومات تمركز ثلاثية البعد للهدف. تنقل هذه المعلومات بواسطة آلات التصوير كمداخل أربعة إلى الشبكة العصبونية. الشبكة هي تصالب ثلاثي الأبعاد من الوحدات التـــي يجب أن تتعلم تطبيق مداخل الكاميرا رباعية البعد إلى وحدة الاستجابة المناسبة. تكون الوحدة بقيم شعاع الأوزان هي الأقرب لشعاع الدخل المسؤولة عن كل نقاط الهدف في منطقة الجوار لهذه الوحدة. تستجيب هذه الوحدة بواسطة إرسال خرج شعاعي ثلاثي البعد ($\Phi,\Theta,\Psi$) لمحركات قيادة الإنسان الآلي لتوضع المقبض عند مكان الهدف. من الواضح أن، التطبيق غير خطي، لأن ثلاثة توضعات إحداثية مختلفة وتمثيلين إحداثيين ستكون محققة.

خلال طور التدريب، اختيرت أماكن الهدف عشوائياً. بعدئذ، يراقب توضع الهدف بواسطة آلات التصوير ويغذى إلى شبكة خريطة الملامح الذاتية التنظيم. ستكون الوحدة ذات شعاع الوزن $w_r$ الذي يكون قريباً جداً من شعاع دخل آلة التصوير هي الوحدة المتكيفة (وحدات الجوار أيضاً تكون متكيفة) وترسل خرجها إلى محركات القيادة.

في البدء، لن تضع شبكة غير مدربة الذراع في المكان الصحيح (مكان الهدف)، بعدئذ تستعمل الشبكة شعاع خطأ الذراع إلى الهدف لتحسين استجابتها. وهذا يمكن أن ينجز بقاعدة تحديث الوزن المبنية على تدرج الهبوط لتخفيف الأخطاء. يجب أن تكتشف الشبكة علاقة التطبيق بأسلوب مستقل وبدون معلم.

النظام كما وصف هنا بسيط جداً وشائق أيضاً. لن تستجيب الشبكة بدقة عالية عندما يكون التطبيق من فراغ مستمر رباعي البعد إلى فراغ متقطع ثلاثي البعد. يتطلب توضّع القابض تماماً على الهدف أن تتعلم الشبكة مسافات الإزاحة بالإضافة إلى اختيار الوحدة المسؤولة. بعد التوضع الأولي للذراع، يلزم تعزيز أكثر كاستشعار خطأ هدف إلى قابض. هذا أيضاً يفرض مقدرات تعليم إضافية على الشبكة وخوارزميتها. اقترح نموذج معزز من قبل Ritter عام 1992[144]. ولكن الوصف الكامل للإنسان الآلي ومشاكله يقع خارج

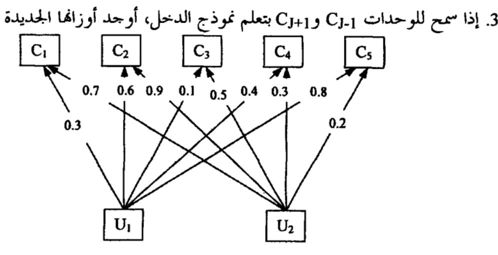
مجال اهتمامنا في هذا الكتاب.

## 12.13 تمارين

1.13 أثبت أنه إذا كانت أشعة الأوزان ليست بطول متساوٍ، فمن الممكن أن شعاع الوزن الذي يظهر الأقرب إلى شعاع الدخل لن يكون شعاع الوزن الذي اختير عند استعمال مسافة الجداء النقطي. وبمعنـــى أدق، ليكن لدينا شعاعا الوزن $\mathbf{w}_1$ و$\mathbf{w}_2$ بطول $\|\mathbf{w}_1\|$ و $\|\mathbf{w}_2\|$ وبزوايا مع المحور الأفقي $\theta_1$ و $\theta_2$ على الترتيب. وإذا كان شعاع الدخل هو $\mathbf{s}$، فما هي المتراجحة (بحدود الأطوال والزوايا) التـــي تعين العصبون الممثل بـــ $\mathbf{w}_1$ ليكون مختاراً كرابح (باستعمال مسافة الجداء النقطي). أعط مثالاً يبين أين لا يكون هذا الاختيار مرغوباً.

2.13 لدينا شبكة خريطة الملامح الذاتية التنظيم المبينة في الشكل (52.13).

1. استعمل مربع المسافة الإقليدية لإيجاد وحدة التجمع $C_J$ التـــي تكون هي الأقرب إلى شعاع الدخل (0.2 , 0.5).
2. استعمل معدل تعليم يساوي 0.2، أوجد الأوزان الجديدة للوحدة $C_J$.
3. إذا سمح للوحدات $C_{J-1}$ و$C_{J+1}$ بتعلم نموذج الدخل، أوجد أوزانها الجديدة




**الشكل 52.13:** شبكة خريطة الملامح الذاتية التنظيم

3.13 كرر التمرين السابق في حالة شعاع الدخل (0.5, 0.5) و$\alpha = 0.1$.

4.13 ليكن لدينا شبكة بوحدتين في طبقة التجمع وخمس وحدات في طبقة الدخل. وأشعة أوزان وحدات التجمع هي :

$\mathbf{w}_1 = (1.0, 0.8, 0.6, 0.4, 0.2)$

$\mathbf{w}_2 = (0.2, 0.4, 0.6, 0.8, 1.0)$

استعمل مربع المسافة الإقليدية لإيجاد وحدة التجمع الرابحة لنموذج الدخل

$\mathbf{x} = (0.5, 1.0, 0.5, 0.0, 0.0)$

باستعمال معدل تعليم $\alpha = 0.2$، أوجد الأوزان الجديدة للوحدة الرابحة.

5.13 ليكن لدينا شبكة تعليم التكميم الشعاعي LVQ ذات وحدتي دخل وأربعة صفوف هدف منشود $C_1$، $C_2$، $C_3$، $C_4$. هناك 16 وحدة تصنيف، بأشعة وزن مشار إليها بالإحداثيات التالية، اقرأ وفق ترتيب سطر عمود. مثلاً، الوحدة ذات شعاع الوزن (0.2,0.4) تخصص لتمثيل الصف الثالث، ووحدات التصنيف للصف الأول لها أشعة أوزان أولية: (0.2, 0.2), (0.2, 0.6), (0.6, 0.8), ( 0.6, 0.4).

| $x_2$ | | | | | | |
|---|---|---|---|---|---|---|
| 1.0 | | | | | | |
| 0.8 | | $C_3$ | $C_4$ | $C_1$ | $C_2$ | |
| 0.6 | | $C_1$ | $C_2$ | $C_3$ | $C_4$ | |
| 0.4 | | $C_3$ | $C_4$ | $C_1$ | $C_2$ | |
| 0.2 | | $C_1$ | $C_2$ | $C_3$ | $C_4$ | |
| 0.0 | | | | | | |
| | 0.0 | 0.2 | 0.4 | 0.6 | 0.8 | 1.0 $x_1$ |

باستعمال مربع المسافة الإقليدية، حدد التغيرات التي تحدث عندما:

1. يكون شعاع الدخل (0.25, 0.25) ممثلاً للصف 1. باستعمال معدل تعليم $\alpha = 0.5$، أثبت أي وحدة صف ستتحرك. أي عين شعاع وزنها الجديد.
2. يكون شعاع الدخل (0.4, 0.35) ممثلاً للصف 1، ماذا سيحدث؟
3. عوضاً عن تمثيل الشعاع الثاني كما في الطلب السابق، قدم الشعاع (0.4, 0.45)، ماذا يحدث؟
4. لنفترض أن دخل التدريب مستنبط من المناطق التالية :

الصف الأول: $0.0 \le x_2 < 0.5 \quad 0.0 \le x_1 < 0.5$

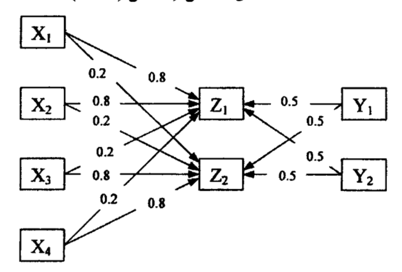
الصف الثانـــي: $0.0 \le x_2 < 0.5 \quad 0.5 \le x_1 < 1.0$

الصف الثالث : $0.5 \le x_2 < 1.0 \quad 0.0 \le x_1 < 0.5$

الصف الرابع : $0.5 \le x_2 < 1.0 \quad 0.5 \le x < 1.0$

من وجهة نظر الدور القصير، هل الأشعة المقدمة في الطلب 2 (0.4, 0.35) أو في الطلب الثالث (0.4, 0.45) لها تأثير أفضل في تحريك وحدات التصنيف باتجاه الأماكن المرغوب بها لتمثيل معطيات الدخل؟.

6.13 ليكن لدينا شبكة الانتشار المتعاكس الكامل (الشكل (53.13) التالية:




**الشكل 53.13**: شبكة الانتشار المتعاكس الكامل

باستعمال زوج التدريب:

$$\mathbf{y} = (1, 0) \qquad \mathbf{x} = (1, 0, 0, 0)$$

نفذ الطور الأول للتدريب (الخطوة الأولى فقط). أوجد تفعيل وحدات طبقة التجمع. حدث الأوزان باستعمال معدل تعليم يساوي 0.3.

7.13 كرر التمرين السادس السابق، باستعمال

$$\mathbf{y} = (0, 1) \qquad \mathbf{x} = (1, 0, 1, 1)$$

8.13 عدل التمرين السادس باستعمال شبكة الانتشار المتعاكس الأمامية فقط.

9.13 صمم شبكة الانتشار المتعاكس لحل مسألة ترافق مجموعة من الأشعة **x** الثنائية (ذات بعد يساوي 6) مع أشعة **y** ثنائية مناسبة (بمركبتين) معرفة كما يلي :

إذا كان اثنان أو أكثر من أول ثلاث مركبات من الشعاع **x** بقيمة 1، عندئذ المركبة الأولى للشعاع **y** ستكون 1 (وإلا ستكون صفراً). وبالمثل، إذا كان اثنان أو أكثر من آخر ثلاث مركبات من الشعاع **x** بقيمة 1، عندئذ المركبة الثانية للشعاع **y** ستكون 1 (وإلا ستكون صفراً).

ناقش كيف نختار عدد وحدات طبقة التجمع (طبقة Kohonen). وأعط وصفاً لعملية تدريب الشبكة. وضح العملية في حالة الزوج

$$(1, 0) \leftrightarrow (1, 1, 0, 0, 1, 0)$$

10.13 أثبت أن أشعة الدخل ذات البعد n يمكن أن تقلب إلى معيارية ببعد n + 1 بواسطة العملية التالية:

إيجاد N بحيث $N > \|v\|$ لكل v. في حالة شعاع مركبته رقم (n + 1) تساوي $\left(N^2 - \|V\|^2\right)^{1/2}$ .أثبت أيضاً أن الشعاع المراد له نظيم يساوي N.

11.13 أثبت أن شبكتـــي الانتشار المتعاكس الكامل والأمامي فقط متكافئتان إذا سلسلت أزواج دخل التدريب **(x, y)** في الانتشار المتعاكس الكامل وعوملت الأشعة المسلسلة كدخل تدريب ونموذج منشود في شبكة الانتشار المتعاكس الأمامية فقط.

12.13 اكتب برنامجاً لأداء الشبكة العصبونية خريطة الملامح الذاتية التنظيم. استعمل وحدتـــي دخل، و50 وحدة تجمع، وبنية طبولوجية لوحدات التجمع. اسمح للوحدة الفائزة بالمنافسة ولجاراتها في منطقة الجوار الطبولوجي الأقرب بالتعلم (بعبارة أخرى، إذا كانت الوحدة J هي الفائزة، عندئذ جاراتها J–1 وJ+1 أيضاً ستتعلم، ما لم يكن J–1<1 أو J+1>50).

استعمل معدل تعليم أولـــي يساوي 0.5، ومن ثم قم بتخفيضه تدريجياً وخطياً إلى 0.01 (عبر 100 دور). ستوضع الأوزان الأولية لوحدات التجمع كافة عشوائية بين (–1 و+1) (لكل مركبة شعاع وزن ولكل وحدة).

ولد ملف معطيات التدريب كما يلي :

اختر رقمين عشوائيين بين (–0.5 و0.5)، وأطلق عليهما اسم $x_2, x_1$. ضع النقطة $(x_1,x_2)$ في مجموعة التدريب إذا كان:

$$x_1^2 + x_2^2 < 0.25$$

كرر حتـــى يكون لديك 100 نقطة تدريب. بعد كل 10 أدوار تدريب، خطط وحدات التجمع (باستعمال شعاع وزنها كمكان في المستوي الإقليدي ثنائي البعد)؛ ارسم خط الوصل من $C_1$ إلى $C_2$، ومن $C_2$ إلى $C_3$،..الخ. لترى علاقاتها الطبولوجية. ستبدأ مع مجموعة مبعثرة حقيقية للتوضع الأولي للأوزان، التـــي ستتحسن تدريجياً لتعطي خطاً يشق طريقه عبر المنطقة التـــي اختيرت نقاط التدريب منها.

13.13 اكتب برنامجاً لأداء الشبكة LVQ الموصوفة في التمرين الخامس. درب الشبكة بالمعطيات المعطاة في نفس التمرين السابق. نفذ باستعمال معدلات تعليم مختلفة، وعدد مختلف من وحدات التصنيف، وهندسات مختلفة لمعطيات الدخل، و. .الخ.

14.13 كرر المثال الهندسي رقم 3، باستعمال ترتيب دخل عشوائي لنقاط المعطيات وباستعمال أول خمس نقاط تدريب من كل صف لوضع الأوزان بقيم أولية لأشعة التجمع الخمس لذاك الصف.

15.13 كرر المثال الهندسي رقم 3 باستعمال LVQ2، وكرر نفس المثال ثانية باستعمال LVQ2.1.

16.13 اكتب برنامجاً لأداء خوارزمية الانتشار المتعاكس بـــ 63 وحدة دخل، و26 وحدة في طبقة التجمع، و15 وحدة في الطبقة Y. اقرأ الأوزان الأولية من الملف. جميع المعطيات موضحة في الأشكال (54.13)، (56.13).

1. في طور التدريب، استعمل التخفيض الخطي التالي لمعدلات التعليم:

0.9 ,0.8 ,0.7 ,... ,0.2 ,0.1. أدخل أزواج أشعة التدريب من الملف. خزن الأوزان النهائية في الملف أيضاً.

2. في طور الاختبار، اطبع دخل نموذج الاختبار، وعين تقريب زوج النماذج. استعمل المداخل التـــي توافق كل الأشعة **x** المستعملة في التدريب (بأصفار للمركبات الموافقة لـــ **y**)، والمداخل التـــي توافق كل الأشعة **y** المستعملة في التدريب (بأصفار للمركبات الموافقة لـــ **x**)، والنسخ الضجيجية لنماذج التدريب.

3. حاول باستعمال مسافة الجداء النقطي، مع مداخل معيارية بطول يساوي الواحد (المسافة الإقليدية). كرر، باستعمال مسافة هامنغ لمعايرة الأشعة.

الدخل من التشكيلة الأولى

```
00##000            ######0            00#####
000#000            0#0000#            0#0000#
000#000            0#0000#            #000000
00#0#00    0#0     0#0000#    ##0     #000000    ###
00#0#00    #0#     0#####0    #0#     #000000    #00
0#####0    ###     0#0000#    ##0     #000000    #00
0#000#0    #0#     0#0000#    #0#     #000000    #00
0#000#0    #0#     0#0000#    ##0     0#0000#    ###
###0###            ######0            00####0
A.....             .B....             ..C...
```

```
#####00            #######            000####            ###00##
0#000#0            0#0000#            00000#0            0#00#00
0#0000#    ##0     0#00000    ###     00000#0    00#     0#0#000    #0#
0#0000#    #0#     0#0#000    #00     00000#0    00#     0##0000    ##0
0#0000#    #0#     0###000    ###     00000#0    00#     0##0000    #00
0#0000#    #0#     0#0#000    #00     00000#0    #0#     0#0#000    ##0
0#0000#    ##0     0#00000    ###     0#000#0    0#0     0#00#00    #0#
0#000#0            0#0000#            0#000#0            0#000#0
#####00            #######            00###00            ###00##
...D...            ....E..            ......J.           ......K
```

الشكل 54.13: معطيات التدريب

الدخل من التشكيلة الثانية

```
000#000            ######0            00###00
000#000            #00000#            0#000#0
000#000    0#0     #00000#    ##0     #00000#    ###
00#0#00    #0#     #00000#    #0#     #000000    #00
00#0#00    ###     #####0     ##0     #000000    #00
0#000#0    #0#     #00000#    #0#     #000000    #00
0#####0    #0#     #00000#    ##0     #00000#    ###
0#000#0            #00000#            0#000#0
0#000#0            ######0            00###00
A.....             .B....             ..C...
```

```
#####00            #######            00000#0            #0000#0
#0000#0            #000000            00000#0            #000#00
#00000#    ##0     #000000    ###     00000#0    00#     #00#000    #0#
#00000#    #0#     #000000    #00     00000#0    00#     #0#0000    ##0
#00000#    #0#     #####00    ###     00000#0    00#     ##00000    #00
#00000#    #0#     #000000    #00     00000#0    #0#     #0#0000    ##0
#00000#    ##0     #000000    ###     0#000#0    0#0     #00#000    #0#
#0000#0            #000000            0#000#0            #000#00
#####00            #######            00###00            #0000#0
...D...            ....E..            ......J.           ......K
```

الشكل 55.13: معطيات التدريب

الدخل من التشكيلة الثالثة:.

```
000#000          ######0          00###0#
000#000          0#0000#          0#000##
00#0#00   0#0    0#0000#   ##0    #00000#   ###
00#0#00   #0#    0#0000#   #0#    #000000   #00
0#000#0   ###    0#####0   ##0    #000000   #00
0#####0   #0#    0#0000#   #0#    #000000   #00
#00000#   #0#    0#0000#   ##0    #00000#   ###
#00000#          0#0000#          0#000#0
##000##          ######0          00###00
A.....           .B....           ..C...
```

```
#####00          #######          0000###          ###00##
0#000#0          0#0000#          00000#0          0#000#0
0#0000#   ##0    0#00#00   ###    00000#0   00#    0#00#00   #0#
0#0000#   #0#    0####00   #00    00000#0   00#    0#0#000   ##0
0#0000#   #0#    0#00#00   ###    00000#0   00#    0##0000   #00
0#0000#   #0#    0#00000   #00    00000#0   #0#    0#0#000   ##0
0#0000#   ##0    0#00000   ###    00000#0   0#0    0#00#00   #0#
0#000#0          0#0000#          0#000#0          0#000#0
#####00          #######          00###00          ###00##
...D...          ...E..           .....J.          .....K
```

**الشكل 56.13**: معطيات التدريب

17.13 اكتب برنامجاً لأداء خوارزمية الانتشار المتعاكس الكامل للمثال التاسع.

18.13 اكتب برنامجاً لأداء خوارزمية الانتشار المتعاكس الأمامي فقط للمثال العاشر.

19.13 لتكن الأرقام العربية ممثلة بالشكل التالي:

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 0: | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1: | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2: | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 3: | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 4: | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 5: | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 6: | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 7: | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

استعمل شبكة الانتشار المتعاكس (الأمامي أو الكامل) لتطبيق هذه الخانات إلى تمثيلاتها (رموزها)الثنائية التالية :

| | | | |
|---|---|---|---|
| 0: | 0 | 0 | 0 |
| 1: | 1 | 0 | 0 |
| 2: | 0 | 1 | 0 |
| 3: | 1 | 1 | 0 |
| 4: | 0 | 0 | 1 |
| 5: | 1 | 0 | 1 |
| 6: | 0 | 1 | 1 |
| 7: | 1 | 1 | 1 |

1. استعمل المسافة الإقليدية

2. كرر في حالة مسافة الجداء النقطي، بمداخل ومخارج منشودة معيارية.

20.13 استعمل الانتشار المتعاكس لحل مسألة المثال الهندسي 2 وقارن النتائج مع ماسبق من نتائج.

# نظرية الطنين المتكيف

# Adaptive Resonance Theory

في هذا الفصل سننظر في صنف آخر هام من الشبكات التكرارية تدعى شبكات نظرية الطنين المتكيف ART (Adaptive Resonance Theory). درس Stephen Grossberg وزملاؤه هذه الشبكات، وطوِّرت بتوسُّع، ودرسها أيضاً Gail Carpenter وأعضاء مجموعة بحث مركز الأنظمة المتكيفة. وتعتبر جامعة Boston أول من اقترح ودرس بنـــى نظرية الطنين المتكيف الأولية منذ منتصف السبعينيات حتـــي الثمانينيات. منذ ذلك التاريخ، عممت نظرية الطنين المتكيف، ودرست دراسة موسعة، واستُعملت في مجال واسع من التطبيقات.

سنبدأ هذا الفصل بفقرة تمهيدية ثم ننتقل إلى وصف النظرية الأساسية، وعمل شبكات الصنف الأبسط لنظرية الطنين المتكيف والمسماة ART1، بعدئذ، سنبحث في النسخة المعممة للنسخة ART1 والمسماة شبكات ART2، وخوارزميات التعليم الموافقة لكلا نوعي الشبكات. وسننظر في بعض التعميمات. وأخيراً، سنصف بعض التطبيقات.

## 1.14 تمهيد

في تطور نماذج الشبكات العصبونية للأنظمة البيولوجية، أصبح من المتوقع أن بعض الخواص الأساسية للشبكات الصنعية تصلح للمقارنة مع مثيلاتها في الشبكات البيولوجية. إن ما نرغب به هو أن تكون شبكاتنا الصنعية قادرة على التكيف المستمر في الوسط المحيط المتغير. وهذا يعنـــي، أن تكون قادرة على الاحتفاظ بالحقائق المفيدة والمعلومات في ذاكرتها بنفس الوقت الذي تتعلم فيه حقائق جديدة هامة.

إذاً، يجب أن لا تمحي الحقائق الجديدة المتعلمة المعلومات المفيدة المكتسبة من قبل. بنفس الوقت، نرغب أن تكون نماذج شبكاتنا الصنعية قادرة على تجاهل المعلومات التـــي ليس لها علاقة بالموضوع وحتـــي عليها أن تنسى المعلومات غير المفيدة والتـــي لا أهمية لا. إذاً، سيكون هناك تعلم مستمر وتخزين للحقائق الهامة المفيدة مع تناسي كل ما لا ينفع من المعلومات.

بكلمات أخرى، نرغب في أن تبدي شبكاتنا العصبونية، التـــي صرفنا في دراستها وتطويرها منذ الأربعينيات وحتـــي اليوم الجهد الكبير، درجة عالية من الاستقرارية عند تعلمها بأسلوب متكيف لمفاهيم وفئات جديدة. لا نريد من الشبكة نسيان (أو ضياع) أو حتـــي تبديل الحقائق المفيدة المخزنة في ذاكرتها من قبل لكي تستطيع أن توافق وتوازن بينها وبين المعارف الجديدة المكتسبة.

من ناحية أخرى، نريد من شبكاتنا أن تكون قابلة للتكيف، ومرنة (مطاوعة وليست عنيدة) كفاية لتكون قادرة على التمييز بين المعلومات المفيدة وغير المفيدة (التـــي ليس لها صلة بالموضوع المدروس).

هاتان صفتان متعارضتان؛ الاستقرارية والمرونة، وهما اللتان دعاهما Grossberg .بمعضلة الاستقرارية مع اللدونة (stability-plasticity). إذ كيف تستطيع الشبكة الاحتفاظ باستقرارها مع بقائها مرنة كفاية لتتكيف على نحو نافع في وسط محيط متقلب الظروف والأحوال؟!.

نأمل أيضاً من الشبكة أن تكون أنيقة وألا تحتاج إلى أن تكون ضخمة بقدر غير ملائم للاحتفاظ بكلتا مميزتـــي الاستقراية واللدونة خلال التكيف عبر دور ممتد من الزمن.

طورت نظرية الطنين المتكيف خلال مدة طويلة واعتبرت توسعة لأنظمة التعليم المتنافس/التعاوني. وكان هذا التطوير محاولة للتغلب على مشكلة الاستقرارية مع اللدونة ومميزات التعليم غير المستقر الأخرى المرافقة للشبكات التنافسية.

امتازت الشبكات الناتجة بعدد من السمات الهامة التـــي تفتقدها بنـــى الشبكات العصبونية الأخرى؛ بما في ذلك التعلم في الزمن الحقيقي الفوري (on–line)، ومقدرتها على التنظيم الذاتـــي (تعلم بدون معلم)، وذاكرة ذاتية الاستقرار بالاستجابة لنماذج دخل عديدة

كيفية، والبحث المتكيف السريع لضبط أفضل للنماذج من مرحلة الدخل إلى مرحلة التخزين، ومقدرة تعلم سريعة (أو بطيئة)، ورفض نماذج الدخل غير المألوفة عندما تصل الشبكة إلى سعة ذاكرتها (للإشباع) المحددة، ومعيار خطأ متغير يسمح بتنظيم متغير لمجموعات الفئات، واحتفاظ ناجح بخواص الاستقرارية مع اللدونة خلال حياة عمل النظام.

بالطبع هناك بعض المساوئ الصغيرة نسبياً وكذلك القيود التـــي تحد من فعاليتها كتعقيد عام في الشبكة، واقتصارها في العمل على الأنظمة الثنائية فقط (شبكات ART1)، وصعوبة في وضع وسيط معيار الخطأ المناسب لبعض التطبيقات، واستعمال غير فعال نسبياً لعصبونات الخرج (يلزم عصبون واحد لكل فئة متعلمة).

تقوم شبكات الطنين المتكيف بتطبيق نماذج الدخل ذات البعد n إلى فئات أو صفوف الخرج المشكلة من ملامح نموذج الدخل. تقسم نماذج الدخل المتشابهة (الجوار الأقرب) إلى مجموعات بنفس الصف والنماذج غير المتشابه إلى صفوف منفصلة متمايزة. يجري تعديل درجة التشابه اللازمة لمجموعات النماذج داخل الصف بحيث تُحدَث عدة مجموعات صف ذات نماذج متشابهة عندما توضع قيمة عتبة التشابه بمستوى عالٍ. في الطرف المقابل، يجري إحداث صفوف أقل عندما توضع قيمة منخفضة للعتبة. في الحالة الأخيرة، تعالج أعضاء الصف بدرجة منخفضة من التشابهية. يمكن أن توضَع قيمة العتبة يدوياً أو ديناميكياً نتيجة عمل الشبكة المعتمد على المسألة المعالجة، وهذا ما يسمح لشبكة نظرية الطنين المتكيف بأن تكون اختيارية أكثر أو أقل لمجموعات النماذج كلما استمرت في التكيف.

يحدث التعليم في شبكات نظرية الطنين المتكيف طبيعياً في الزمن الحقيقي خلال العمل العادي للشبكة. وهو نوع من التعليم المتكيف المستمر بدون معلم، حيث تكوَّن فئة جديدة آلياً عندما يقدَّم نموذج جديد إلى الشبكة. يستمر تكوين الفئات الجديدة من المداخل الجديدة حتـــى تستنْزف الشبكة حوضها بالكامل (جميع عقد الخرج) من عصبونات فئات الخرج غير المستخدمة من قبل، عند ذلك ترفض الشبكة أي دخل جديد آخر.

تنظم نماذج الدخل التـــي تكون مشابهة للفئات المنشأة من قبل دون إبطاء بإعطاء خرج عالٍ عندما يختار عصبون الفئة. أيضاً، تبدأ المداخل المنسجمة مع إحدى الفئات بدرجة ما من التعليم للفئة المعطاة، وبنفس الوقت، دون إفساد استقرارية الفئات المتعلمة من قبل.

سنصف أولاً، عمل وديناميكية أبسط نسخ شبكات نظرية الطنين المتكيف، ART1،

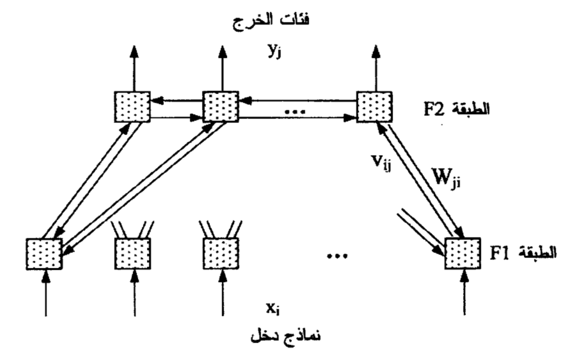
وبعدئذ، سنواصل رحلتنا لتعميم الشبكة الأساسية، ART2، بما في ذلك، تغيرات أخرى على بنيات الشبكة.

## 2.14 بنية شبكة نظرية الطنين المتكيف الأولى

**Architecture ( ART1) The Adaptive Resonance Theory Network**

صممت شبكة نظرية الطنين المتكيف الأولى في حالة مداخل ثنائية فقط. تنجز هذه الشبكات تطبيقاً من نماذج الدخل الثنائية ببعد n إلى فئة الخرج الوحيدة.

تتألف الشبكة من طبقتين متصلتين اتصالاً كاملاً استناداً إلى أوزان (أدنـــى لأعلى) معدلة (متكيفة) على كل الوصلات من عقد الطبقة الدنيا، الطبقة F1، إلى عقد الطبقة العليا (طبقة التجمع)، الطبقة F2، وتكون الأوزان ( أعلى لأدنى) متكيفة أيضاً على كل وصلات التغذية العكسية الواصلة من عقد الطبقة العليا عكسياً إلى عقد الطبقة الدنيا. إضافة إلى ذلك، هناك وصلات بين كلا عصبونات الطبقة F1 وF2 إلى عصبونات خاصة تنجز وظائف التحكم (سنشرح ذلك فيما بعد). البنية الأساسية لهذه الشبكة موضحة في (الشكل 1.14).




**الشكل 1.14**: بنية شبكة نظرية الطنين المتكيف الأولى

عندما يقدم نموذج الدخل إلى عصبونات الطبقة F1 يولد تفعيل الخرج بواسطة كل

العصبونات. بعدئذ، ترسل إشارة خرج الطبقة F1 إلى الطبقة F2 نتيجة الأوزان (أدنـــى لأعلى) $v_{ij}$. تربح عصبونات الطبقة F2 ذات الانسجام الأقرب بين شعاع الإشارة الواصل من الطبقة F1 وشعاع وزنها الموافق V منافسة "الرابح يحوز الكل" من بين عصبونات الطبقة F2. بعدئذ، تنتج عصبونات الطبقة F2 الرابحة إشارة تفعيل خرجها؛ المسمى توقع ذاكرة أعلى لأدنـــى المخزن، والذي ينتشر عكسياً إلى الطبقة F1 نتيجة أوزان التغذية العكسية $w_{ji}$ .

تقاد مخارج كل عصبونات الطبقة F2 الأخرى إلى الصفر في عملية المنافسة وستمنع من إعطاء إشارات خرج. تقارن الإشارة المثقلة المغذاة عكسياً إلى الطبقة F1 مع إشارة الدخل عند الطبقة F1. إذا كان الانسجام قريباً جداً بين الإشارتين، تجري تقوية تفعيل خرج الطبقة F1 (لاحظ أن الإشارة المغذاة عكسياً بواسطة عصبون F2 الرابح للمنافسة يمكن أن تغير مستويات تفعيل عصبونات F1). تعزز إشارة خرج F1 المقواة الإشارة المغذاة عكسياً من عصبون F2 الرابح، وينتج نوع من الطنين بين الطبقتين. عندما تستقر هذه العملية، تعطي عصبونات الطبقة F2 الرابحة إشارة خرج عالية لتشير إلى الفئة المختارة للنموذج المقدم إلى الدخل.

عندما لا يكون الانسجام بين نموذج الدخل والإشارة المستدعاة من الطبقة F2 قريباً كفاية، فإن إشارة تصفير (reset) تجبر عصبون F2 ليكون غير فعال خلال حياة الدور. بعد التصفير، تفعل ثانية إشارة الدخل عند الطبقة F1 من جديد، ويمكن لعصبون آخر في الطبقة F2 بعدئذ أن يصبح الرابح في منافسة الرابح يحوز الكل ثانية.

إذا لم يكن الانسجام بين نموذج الدخل ونموذج إشارة التغذية العكسية من الرابح F2 الجديد قريباً كفاية ثانية، يحدث تصفير آخر وعدم تفعيل للرابح F2 الثاني.

تستمر عملية الاختبار الشرطية هذه حتـــي يوجد انسجام جيد أو حتـــي تصبح كل العصبونات المستخدمة في F2 (الحوض) غير فعالة (لم يبق أي عقدة غير مستخدمة). في الحالة الأخيرة، يختار عصبون جديد غير مستخدم من قبل ليصبح فئة محدثة من جديد. أحدثت الفئة الجديدة بوضع أوزان وصلات التغذية الأمامية والعكسية بنفس قيم النموذج كنموذج الدخل الثنائي.

لدى إعطاء نموذج دخل جديد، إذا كانت كل العصبونات المتوفرة في الطبقة F2

مستخدمة سابقاً في الفئات ولم يوجد الانسجام المقبول من أدنـــى لأعلى ومن أعلى لأدنـــى (وهذا الانسجام هو ما نعنـــي به الطنين بين الطبقتين F1 وF2)، عندها سيُرفَض نموذج الدخل، ولن يكون أي عصبون خرج في F2 فعالاً، ولن يحدث تعليم.

في الواقع، يحدث التعليم المتكيف فقط عندما يوجد انسجام جيد وبعدها سينشأ الطنين (بالطبع، يحدث تعلم نموذج مباشرة فيما إذا اختير عصبون مستخدم من جديد). ينجز التعليم من خلال تعديل الأوزان في كلا ممرات التغذية الأمامية (أدنـــى لأعلى) وممرات التغذية العكسية (أعلى لأدنى) وذلك لإزاحة القيم في اتجاه نموذج الدخل ؛ أي النموذج الذي كان سابقاً قريباً من نموذج الفئة الأولي المخزن.

لقد أطلق اسم نظرية الطنين المتكيف على هذه الشبكات للدلالة على أن التعليم المتكيف العادي يحدث فقط خلال الطنين بين الطبقتين F1 وF2. إن شبكات نظرية الطنين المتكيف مزودة بشبكتين جزئيتين إضافيتين لإنجاز وظائف التحكم.

تدعى الشبكة الجزئية الأولى بتصفير ذاكرة الأجل القصير STM reset؛ (Short Term Memory reset)، وهي جزء من نظام توجيه جزئي (أشير له بـــ A في (الشكل 2.14). ترسل هذه الشبكة الجزئية الأولى STM reset إشارة لمنع تفعيل عقدة طبقة F2 عندما لا يكون الانسجام بين نموذج الدخل ونموذج تفعيل الطبقة F2 قريباً كفاية. تعتمد درجة عدم الانسجام المسموح بها، قبل توليد إشارة التصفير، على قيمة العتبة القابلة للتعديل $\rho$ والمسماة بوسيط الاحتراس (أو اليقظة) (vigilance parameter).

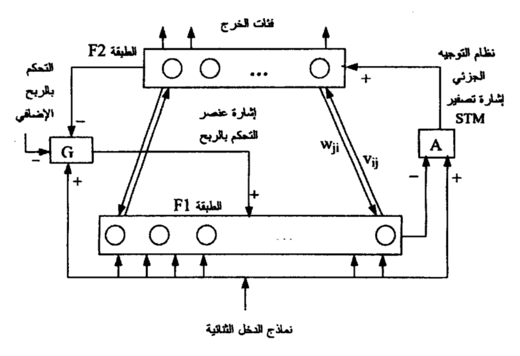
تدعى الشبكة الجزئية الثانية عنصر تحكم ربح الانتباه، كما هو موضح في (الشـــكل 2.14). وظيفة هذه الشبكة الجزئية تنظيم عملية اختبار شروط الفئة والاستقرار الذاتـــي للتعليم. فهي توفر الميكانيكية اللازمة التـــي تستطيع الطبقة F1 بواسطتها التفريق بين إشارات أدنـــى لأعلى وأعلى لأدنـــى المستقبلة.

هذه الشبكة الجزئية لها ثلاثة مداخل: إشارة دخل أدنـــى لأعلى مهيجة، وإشارة توقع أعلى لأدنـــى مخمدة وإشارة ضمن النموذج (إشارة التحكم بفتح وإغلاق نموذج الدخل). تصبح شبكة التحكم بالربح الإضافي فعالة، وتعطي خرجاً للعقدة F1، عندما يقدم نموذج دخل I للطبقة F1. يهيج هذا الخرج بالتساوي كل عقدة في الطبقة F1 سامحاً للعقد F1 أن

تصبح فعالة كفاية لإرسال إشارات خرجها إلى الطبقة F2.

عندما تصبح الطبقة F2 فعالة، تغلق شبكة التحكم بالربح بواسطة إشارة مخمدة من F2، ومن ثم، يجب على العقد F1 أن تستقبل إشارة دخل (توقع) معززة من الطبقة F2 لكي تؤازر الخرج.

بكلمات أخرى، يجب أن تستقبل الطبقة F1 اثنين من مداخلها الثلاثة الفعالة لتبقى فعالة، وإلا سيتناقص الخرج الكلي من F1. الحاجة لأن يكون اثنان من ثلاثة مداخل فعالة ليكون الخرج متولداً تعرف بقاعدة 3/2. من الواضح أن هذه الشبكة الجزئية تتفاعل مع عملية معالجة تصفير ذاكرة الأجل القصير، وسنصف عملية الشبكة الجزئية بالتفصيل فيما بعد.




**الشكل 2.14**: ممرات تدفق إشارة التحكم والنموذج في شبكة نظرية الطنين المتكيف الأولى

## 3.14 ديناميكات شبكة نظرية الطنين المتكيف الأولى ART1 Dynamics

يشار للأوزان المعدلة في شبكة نظرية الطنين المتكيفة الأولى كذاكرة الأجل الطويل LTM (Long Term Memory)، لأن هذه الأوزان تتغير أو تتكيف ببطء نسبياً خلال الزمن لتمثيل نماذج الدخل. من ناحية أخرى، يشار إلى تفعيلات العصبونات المسببة بواسطة

إشارات الدخل كذاكرة الأجل القصير STM؛ (Short Term Memory)، لأنها تعيش قليلاً ولمدة دور معالجة واحد فقط.

يتضح من الأشكال (1.14) و(2.14) أن الطبقة F1 يشار إليها كلياً كطبقة حقل ملمح العصبونات، ويشار إلى الطبقة F2 إجمالاً كحقل الفئة أو طبقة خرج العصبونات (طبقة التجمع).

لفهم ديناميكية شبكات نظرية الطنين المتكيف الأولى وكيف تتفاعل الطبقات، نحتاج إلى بعض التعاريف.

ليكن **I** شعاع دخل ثنائياً خارجياً، حيث $I = (I_1 ,I_2 , \ldots ,I_n)$، و$I_i$ بت الدخل للعقدة i من الطبقة F1. وليكن $x_i$ قيمة إشارة الفعالية المستقبلة عند العقدة i، $I = 1 ,2\ldots ,n$ (يعطى الشعاع **X** لإشارات الفعالية، والمسمى أثر الأجل القصير للطبقة F1، قيمة أولية مساوية للدخل **I**، ويمكن عموماً، أن تختلف هذه القيمة تبعاً لإشـــارات التغذية العكســـية من الطبقة F2).

يرشح تفعيل الخرج $x_i$ من العقدة i ( يضرب بالمعامل) بـــ $v_{ij}$، وهي مركبة ذاكرة الأجل الطويل (الوزن) على وصلة التغذية الأمامية بين العقدة I في الطبقة F1 والعقدة j في الطبقة F2 لتعطي $f_i = x_i \upsilon_{ij}$ . تمرر هذه الإشارة إلى العقدة j في الطبقة F2 حيث تركب مع إشارات أخرى من F1.

ليكن $S_j$ الإشارة الكلية المستقبلة عند العقدة j للطبقة F2، حيث $j = 1 ,2 ,\ldots ,m$، وتعطى بما يلي:

$$S_j = \sum_{i=1}^{n} x_i \upsilon_{ij} = \sum_{i=1}^{n} R_i \qquad (1.14)$$

وهي إشارة الخرج المتولدة من العقدة j في الطبقة F2 نتيجة التبادل التنافسي فيما بين عقد الطبقة F2.

تربح المنافسة العقدة في الطبقة F2 المستقبلة لأكبر دخل شبكة net. ويعطى خرج العقدة رقم j بواسطة $y_j$، أثر ذاكرة الأجل القصير عند الطبقة F2، بما يلي:

$$y_j = f(S_j) = \begin{cases} 1 & S_j = max\{S_k\} \\ 0 & S_j \neq max\{S_k\} \end{cases} \qquad (2.14)$$

تغذى إشارة تفعيل خرج الطبقة F2، $y = \{y_1, y_2, \dots, y_m\}$، عكسياً إلى عقد الطبقة F1 من خلال أوزان ذاكرة الأجل الطويل أعلى لأدنـــى $w_{ji}$. ترشح هذه الأوزان إشارة تغذية عكسية من العقدة j في الطبقة F2 لتعطي $T_j = y_j w_{ji}$. هذه الإشارات ركبت مع إشارات أخرى من F2 ومررت إلى العقدة i في الطبقة F1. قيمة إشارة التفعيل net المستقبلة عند العقدة i في الطبقة F1 من الطبقة F2 هي $U_I$، وتعطى بـــ :

$$U_i = \sum_{j=1}^{m} y_j w_{ji} = \sum_{j=1}^{m} T_j \qquad (3.14)$$

حيث يُضم الشعاع **U** مع إشارة نموذج الدخل **I** عند F1 ليعطي شعاع خرج جديداً $X^* = \{x_1^*, x_2^*, \dots, x_n^*\}$ من الطبقة F1.

تقارن إشارة الدخل **I** مع إشارة التغذية العكسية **U** في الطبقة F1 لمعرفة هل سيكون انسجامها قريباً أم لا. استعملت المقارنة لتحديد فيما إذا كان التصفير سيعمل أم لا. سيكون التحويل الكامل للإشارات (ذهاباً وإياباً) من الدخل الثنائي المستقبل **I** عند الطبقة F1 إلى استقبال إشارة التغذية العكسية **U** (من F2) عكسياً عند F1 على الشكل التالي:

$$\mathbf{I} \rightarrow \mathbf{X} \rightarrow \mathbf{R} \rightarrow \mathbf{S} \rightarrow \mathbf{Y} \rightarrow \mathbf{T} \rightarrow \mathbf{U}$$

تولد إشارة التصفير من العقدة A لنظام التوجيه الجزئي إذا كان الانسجام بين توقع أعلى لأدنـــى ونموذج الدخل ليس قريباً كفاية. وهذا يحدث عندما يكون دخل الشبكة net لـــ A أقل من قيمة وسيط الاحتراس $\rho$. وتنشأ هذه الحالة عندما تختلف مركبات أعلى لأدنـــى من الطبقة F2 عن قيمة لإشارة الدخل مفروضة من قبل.

سيكون دخل واحد لـــ A مهيجاً (+)، وسوف يتناسب مع إشارة الدخل **I**، وهذا يعني $\pi|\mathbf{I}|$، حيث $|\mathbf{I}|$ يشير إلى عدد المركبات الموجبة (بت واحد) في الدخل **I** و$\pi$ ثابت موجب. وسيكون الدخل الآخر لـــ A إشارة مخمدة (−)، ومتناسباً مع إشارات خرج الطبقة F1، وهذا يعنـــي $\theta|\mathbf{X}|$، حيث يشير $|\mathbf{X}|$ إلى عدد المركبات الموجبة في الشعاع **X** و$\theta$ ثابت موجب.

يجري اختيار الإشارة المخمدة لتكون أكبر من الإشارة المهيجة؛ أي $\theta \leq \pi$، بحيث لا يتولد تصفير عند الطبقة F2 عندما تكون غير فعالة(عندئذ، $|\mathbf{X}|=|\mathbf{I}|$). لذلك يكون وسيط اليقظة $\rho$ معرفاً تعريفاً مناسباً لتكون النسبة معطاة بواسطة $\rho = \pi/\theta \leq 1$ وكي يحدث التصفير عندما $\rho > |\mathbf{X}|/|\mathbf{I}|$، وبالمثل، يمنع معيار انسجام/تصفير إشارة التصفير من الحدوث عندما تزيد نسبة إشارات أعلى لأدنـــى الموجبة إلى الدخل عن العتبة، وهذا يعنـــي أنه عندما:

$$\frac{|\mathbf{X}|}{|\mathbf{I}|} = \frac{|\mathbf{U} \cap \mathbf{I}|}{|\mathbf{I}|} \geq \rho$$

في هذه الحالة، سيحدث الطنين عندما تنسجم العقد الفعالة الخاصة في F2 كفاية مع نفس مركبات إشارة الدخل الفعالة **I** عند F1.

تتطلب قاعدة 3/2 أن يكون اثنين من ثلاثة من مداخل الطبقة F1 فعالة لكي تكون عقد الطبقة F1 فعالة. يرسل أي خرج من F2 إشارة مخمدة إلى عقدة الشبكة الجزئية للتحكم في ربح الانتباه لمنع عقد الطبقة F1 أن تصبح فعالة جداً ما لم تكن إشارات تهييج الانسجام الخاص أيضاً مستقبلة عند الطبقة F1 من الطبقة F2 خلال ممرات ذاكرة الأجل الطويل.

وهكذا، لكي يحدث الطنين، يجب أن تنسجم إشارة أعلى لأدنـــى مع إشارة الدخل للمدى المطلوب بواسطة مستوى عتبة الاحتراس. وبذلك نلاحظ أن معيار 3/2 يسمح للشبكة بالتفريق بين الانسجام وعدمه لإشارات الدخل وتوقع أعلى لأدنى، ويدعم معيار التصفير.

لاحظ أن الدخل الثالث لعقدة التحكم في ربح الانتباه هو إشارة مخمدة (دخل ضمن النموذج) الذي يمنع الدخل **I** فقط من التفعيل الزائد لعقد الطبقة F1.

تضبط فعالية العقدة i في الطبقة F1 بمعادلة الفروق التالية:

$$\delta \frac{dx_i}{dt} = -x_i + (1 - a_1 x_i) J_i^+ - (b_1 + c_1 x_i) J_i^- \qquad (4.14)$$

حيث يعطى دخل التهييج الكلي للعقدة i بما يلي:

$$J_i^+ = I_i + U_i$$

و يعطى دخل التخميد الكلي للعقدة i بـــ:

$$J_i^- = \sum_j f(y_i)$$

(دخل إشارة التحكم في ربح الانتباه الموصوف آنفاً). والوسطاء $\delta$ و$a_1$ و$b_1$ و$c_1$ كلها ليست سالبة.

تعطى معادلات تفعيل عقد الطبقة F2 بواسطة:

$$\delta \frac{dy_i}{dt} = -y_j + (1 - a_2 y_j) J_j^+ - (b_2 + c_2 y_j) J_j^- \qquad (5.14)$$

حيث جميع الوسطاء $\delta$ و$a_2$ و$b_2$ و$c_2$ ليست سالبة، وتعطى إشارة التغذية العكسية الذاتية الموجبة للعقدة j بـــ:

$$J_j^+ = g(y_j) + S_j$$

إشارة الدخل هي مجموع إشارات التغذية العكسية السالبة من كل العقد الأخرى في الطبقة F2، وتعطى بـــ:

$$J_j^- = \sum_{k \neq j} g(y_k)$$

تختار الوسطاء في المعادلات السابقة بحيث تصبح العقدة F2 المستقبلة الدخل net الأكبر $S_j$ هي الرابحة من بين كل عقد F2 غير فاقدة الأهلية. هذا يعني، أن العقدة j تكون رابحة عندما يكون :

$$y_j = f(S_j) = \begin{cases} 1 & S_j = max\{S_k : k \in J\} \\ 0 & S_j \neq max\{S_k : k \in J\} \end{cases} \qquad (6.14)$$

حيث عدلنا المعادلة (2-14) بإضافة دليل عقدة الطبقة F2، وذلك بوضع J الذي يشمل فقط أدلة العقد في الطبقة F2 غير العاجزة (لم تفقد أهليتها) بواسطة إشارة التصفير.

## 4.14 تعليم شبكة نظرية الطنين المتكيف الأولى ART1 learning

يحدث التعليم في ذاكرة الأجل الطويل حينما يوجد الانسجام الكافي ويحدث الطنين، أو حينما تختار عقدة فئة غير مستخدمة من جديد في الطبقة F2. إن المعادلات الضابطة للتعليم في ممرات ذاكرة الأجل الطويل أعلى لأسفل وأسفل لأعلى تتبع قانون Weber وقواعد الاضمحلال المرافقة التـــي تشترط أن يكون لأوزان ذاكرة الأجل الطويل المتعلمة خلال

ترميز نموذج الطبقة F1، X عدد أصغري من المداخل الموجبة لتكون أكبر من أوزان الإشارة X بمركبات موجبة أكثر. هذا الشرط ضروري لنتمكن من تمييز نموذج a من نموذج b بواسطة عقد الفئة F2 عندما تكون a مجموعة جزئية من b ($a \subset b$).

نوقشت تفاصيل قانون Weber وقواعد الاضمحلال المرافقة المستعملة في أداء شبكة نظرية الطنين المتكيف الأولى بالتفصيل من قبل Grossberg عام 1988[19].

يمكن أداء التعليم السريع والبطيء في شبكات نظرية الطنين المتكيف. سنركز فيما يلي على التعليم السريع فقط. في هذه الحالة، تكون معادلات تعليم ذاكرة الأجل الطويل أعلى لأدنـــى/أدنـــى لأعلى نوعاً ما أكثر تعقيداً لتأخذ بعين الاعتبار المنافسة خلال التعليم.

في ممرات أعلى لأدنـــى عدلنا الأوزان وفقاً لـــ:

$$\frac{dv_{ij}}{dt} = kf(S_j)\left[(1-v_{ij})\lambda h(x_i) - v_{ij}\sum_{k\neq j} h(x_k)\right] \qquad (7.14)$$

حيث k و λ ثوابت، و $f(S_j) = y_j$ هو خرج العقدة j في الطبقة F2، و $h(x_j)$ هو الخرج المرسل بواسطة العقدة i في الطبقة F1.

معادلات التعليم أدنـــى لأعلى تكون أبسط نوعاً ما وتعطى بواسطة:

$$\frac{dw_{ji}}{dt} = f(S_j)\left[-w_{ji} + h(x_i)\right] \qquad (8.14)$$

أتت المعادلة (8.14) من المعادلة (7.14) بتبسيط الثوابت المكافئة لـــ k و λ في ((7.14). وللأداء الموصوف هنا، وُضِع كلاهما بقيمة 1 في المعادلة ((8.14). ونتيجة لتطبيق قانون Weber وقاعدة الاضمحلال المرافقة فإن تعليم ذاكرة الأجل الطويل يحدث فقط عندما يوجد الانسجام التام بين نموذج أدنـــى لأعلى وتوقع ذاكرة أعلى لأدنى، أو عندما ينظم نموذج جديد ويتوفر عقد F2 غير مســـتخدمة. يمكن التعبير عن معادلات ذاكرة الأجـــل الطويـــل (7.14) و(8.14) بأشكال مختصرة ستعطى لاحقاً.

لاحظ أن تعليم أدنـــى لأعلى يحدث عندما تكون كل من العقدة i في الطبقة F1 والعقدة j في الطبقة F2 فعالتين، عندئذ سيكون $f(S_j) = y_j = 1$ و $h(x_j) = 1$. من ناحية أخرى، يضمحل $v_{ij}$ بسرعة إلى الصفر عندما تكون العقدة i في الطبقة F1 غير فعالة، لكن

تكون العقدة j في الطبقة F2 فعالة، وأخيراً لا يحدث تعليم في $v_{ij}$ إذا كانت العقدة j في F2 غير فعالة.

- إذا كانت العقدة i في الطبقة F1 والعقدة j في الطبقة F2 فعالتين (1) فإن:

$$\frac{dv_{ij}}{dt} = k[(1-v_{ij})\lambda - v_{ij}(|\mathbf{X}|-1)] \qquad \text{(9.14 أ)}$$

- إذا كانت العقدة i في الطبقة F1 غير فعالة (0) والعقدة j في الطبقة F2 فعالة (1) فإن:

$$\frac{dv_{ij}}{dt} = -k|\mathbf{X}|v_{ij} \qquad \text{(9.14 ب)}$$

- إذا كانت العقدة j في الطبقة F2 غير فعالة (0) فإن:

$$\frac{dv_{ij}}{dt} = 0 \qquad \text{(9.14 ج)}$$

يمكن استنتاج معادلات مشابهة في حالة تعليم ذاكرة الأجل الطويل في الحالات الثلاث المذكورة آنفاً.

وهكذا، يحدث تعليم ما في $w_{ji}$ عندما تكون العقدة i في الطبقة F1 والعقدة j في الطبقة F2 فعالتين معاً (1) فإن $f(S_j) = y_j = 1$ و $h(x_j) = 1$ ويزداد أسياً باتجاه الواحد (العقدة j تحاول تعليم نموذج الفعالية عبر F1)، لكن يضمحل $v_{ij}$ سريعاً (أسياً) إلى الصفر عندما تكون العقدة i في الطبقة F1 غير فعالة، لكن العقدة j في الطبقة F2 تكون فعالة (تؤكد قاعدة 3/2 أن تعليم العقدة i تبقى غير فعالة ما لم يكن هناك دخل داعم I)، ولا يحدث تعليم في $v_{ij}$ إذا كانت العقدة j في F2 غير فعالة، وهكذا:

- إذا كانت العقدة i في الطبقة F1 والعقدة j في الطبقة F2 فعالتين (1) فإن:

$$\frac{dw_{ji}}{dt} = -w_{ji} + 1 \qquad \text{(10.14 أ)}$$

- إذا كانت العقدة i في الطبقة F1 غير فعالة (0) والعقدة j في الطبقة F2 فعالة (1) فإن:

$$\frac{dw_{ji}}{dt} = -w_{ji} \qquad \text{(10.14 ب)}$$

- إذا كانت العقدة j في الطبقة F2 غير فعالة (0) فإن:

$$\frac{dw_{ji}}{dt} = 0 \qquad \text{(ج 10.14)}$$

للتوثق من أن عملية البحث الشرطية تتقدم بأسلوب مرتب في الطبقة F2، وأن عقد F2 غير المستخدمة لن تتعلم من نماذج الدخل ما لم تكن مختارة فعلياً لفئة جديدة، من الضروري تقديم قيم أوزان أدنـــى لأعلى وأعلى لأدنـــى $v_{ij}$ و $w_{ij}$ على الترتيب. يمكن برهان أن القيم الأولية لأوزان أدنـــى لأعلى تحقق المتطلبات التالية:

$$0 < v_{ij}(0) < \frac{\lambda}{\lambda - 1 + m} \qquad (11.14)$$

حيث m عـــدد العقد في الطبقة F2. وهذا معروف بمتراجحة الوصـــول المباشـــر (direct access inequality). تحقق قيم أوزان أعلى لأدنـــى الأولية متراجحة تعليم نموذج المعايرة (template learning inequality) التالية:

$$\frac{b_1 - 1}{d} < w_{ji}(0) \le 1 \qquad (12.14)$$

حيث $b_1$ معرف في المعادلة (4.14)، و d ثابت ضرب موجب لمخارج عقدة F2 فعالة. وكذلك، يمكن برهان أنه خلال عملية التعليم السريع تتقارب قيم الأوزان إلى القيم التالية:

$$\begin{aligned} v_{ij} &\cong \lambda/(\lambda - 1 + |\mathbf{X}|) && i \in \mathbf{X} \\ v_{ij} &= 0 && i \notin \mathbf{X} \end{aligned} \qquad (13.14)$$

$$\begin{aligned} w_{ji} &\cong 1 && i \in \mathbf{X} \\ w_{ji} &= 0 && i \notin \mathbf{X} \end{aligned} \qquad (14.14)$$

لكل تجربة تعليم. أعطى Carpenter & Grossberg عام 1987[238][237] القيم المسموح بها للوسطاء المعرفة من قبل المستعمل، وأعطى Lippmann عام 1987[239] لهذه الوسطاء القيم النموذجية التالية:

| الوسيط | المجال المسموح به | قيمة نموذجية |
|---|---|---|
| $L = \lambda$ | $L = \lambda > 1$ | 2 |
| $\rho$ | $0 < \rho \le 1$ | 0.9 |
| $v_{ij} = b_{ij}$ | $0 < v_{ij}(0) < \frac{\lambda}{\lambda - 1 + n}$ | 1/(1 + n) |
| $w_{ji}$ | $w_{ji}(0) = 1$ | 1 |

ثم برهن Grossberg عام 1988[9] عدداً من النظريات التـــي تثبت أن التعليم لاستجابة قائمة كيفية من نماذج الدخل الثنائية لشبكة نظرية الطنين المتكيف الأولى يكون مستقراً ذاتياً، وأن كل النماذج تصل مباشرة إلى فئاقما بعد استقرار عملية تعلم التمييز.

سنناقش فيما يلي بعض الأمثلة البسيطة لفهم خوارزمية تعليم شبكة نظرية الطنين المتكيف الأولى، وسنحرص على ذكر خطوات الخوارزمية مع الحساب المقابل في المسألة المعالجة.

**مثال 1:**

في هذا المثال البسيط سنوضح عمل شبكة نظرية الطنين المتكيف الأولى لتجميع أربعة أشعة مع أخذ وسيط يقظة منخفض. ستكون قيم وسطاء هذا المثال كما يلي:

$n = 4$: عدد المركبات في شعاع الدخل

$m = 3$: العدد الأعظمي للتجمعات المطلوب تشكيلها

$\rho = 0.4$: وسيط الاحتراس

$L = \lambda = 2$: الوسيط المستعمل في تحديث أوزان أدنـــى لأعلى

$v_{ij}(0) = 1/(1+n)$ : الأوزان الأولية أدنـــى لأعلى (يسمح بنصف القيمة العظمى)

$w_{ji}(0)$ : الأوزان الأولية أعلى لأسفل

في هذا المثال سنستعمل خوارزمية شبكة نظرية الطنين المتكيف الأولى لتجميع أربعة ($n = 4$) من أشعة الدخل **I** الثنائية التالية إلى ثلاثة ($m = 3$) تجمعات:

$$(1, 1, 0, 0) \quad (0, 0, 0, 1) \quad (1, 0, 0, 0) \quad (0, 0, 1, 1)$$

ستكون خطوات الخوارزمية على النحو التالي:

1. وضع القيم الأولية للوسطاء:

$$0 < \rho \le 1 \text{ و } \lambda$$

$$\lambda = 2 \,, \rho = 0.4$$

وضع القيم الأولية للأوزان:

$$0 < v_{ij}(0) < \frac{\lambda}{\lambda - 1 + n} \quad , \quad w_{ji}(0) = 1$$

$$v_{ij}(0) = b_{ij}(0) = 0.2 \,, t_{ji}(0) = 1$$

2. ما دام شرط التوقف غير محقق ، كرر الخطوات من 3 إلى 14.

3. فـــي كـــل دخـــل تدريب، كرر الخطوات من 4 إلى 13، في حالة شعاع الدخل الأول (1, 1, 0, 0)

4. ضع تفعيلات كل عقد الطبقة F2 بقيمة الصفر، ضع تفعيلات كل عقد الطبقة F1 بقيمة شعاع الدخل $\mathbf{I} = (1, 1, 0, 0)$

5. احسب نظيم شعاع الدخل:

$$i = 1, 2, \cdots, n \quad , \quad \sum_i I_i = \| \mathbf{I} \|$$

$$\| \mathbf{I} \| = 2$$

6. حساب تفعيلات كل عقدة في طبقة الدخل F1:

$$\mathbf{X} = \mathbf{I}$$

$$x_i = I_i \qquad , i = 1, 2, \cdots, n$$

7. في كل عقدة F2 ليست مخمدة:

إذا كان $y_j \neq -1$ ، عندئذ $\quad y_j = \sum_{i=1}^{n} v_{ij} x_i \qquad , \quad j = 1, 2, \cdots, m$

حساب دخل كل عقدة في الطبقة F2:

$$y_1 = 0.2(1) + 0.2(1) + 0.2(0) + 0.2(0) = 0.4$$

$$y_2 = 0.2(1) + 0.2(1) + 0.2(0) + 0.2(0) = 0.4$$

$$y_3 = 0.2(1) + 0.2(1) + 0.2(0) + 0.2(0) = 0.4$$

8. مادام التصفير صحيحاً (true)، كرر الخطوات 9-12

9. أوجد J (العقدة الرابحة) بحيث $y_J \geq y_j$ لكل عقد j:

إذا كان $y_j = -1$، عندئذ تكون كل العقد مخمدة والنموذج لا يمكن أن يوضع في تجمع.

لما كان لكل الوحدات الدخل نفسه فإن J = 1.

10. إعادة حساب تفعيلات x للطبقة F1:

$$x_i = I_i w_{Ji}$$

الآن $T_1 = (1, 1, 1, 1)$ ومن ثم $\mathbf{X} = (1, 1, 0, 0)$

11. حساب نظيم الشعاع x:

$$\sum_i x_i = \| \mathbf{X} \|$$

$$\| \mathbf{X} \| = 2$$

12. اختبار شرط التصفير:

إذا كان $|\mathbf{X}|/|\mathbf{I}| < \rho$

فإن $y_J = -1$ (J عقدة مخمدة)، استمر في تنفيذ الخطوة 8 ثانية.

أما إذا كان $|\mathbf{X}|/|\mathbf{I}| \geq \rho$

نفذ الخطوة 13

$$|\mathbf{X}|/|\mathbf{I}| = 2/2 = 1 > 0.4$$

ويكون التصفير خطأ. نفذ الخطوة 13

13. حدث الأوزان في حالة العقدة الرابحة J = 1 (تعليم سريع):

$$v_{iJ}^{new} = \lambda x_1 / (\lambda - 1 + \|\mathbf{X}\|) \quad , \quad W_{Ji}^{new} = x_i$$

تحديث $\mathbf{v}_1$: في حالة L = 2 ستكون الأوزان:

$$v_{iI}^{new} = 2x_i / (1 + \|\mathbf{X}\|)$$

لذا، تصبح مصفوفة وزن أدنـــى لأعلى:

$$\begin{bmatrix} \mathbf{0.67} & 0.2 & 0.2 \\ \mathbf{0.67} & 0.2 & 0.2 \\ \mathbf{0} & 0.2 & 0.2 \\ \mathbf{0} & 0.2 & 0.2 \end{bmatrix}$$

تحديث $\mathbf{w}_1$: قيم وزن التعليم السريع

$$t_{Ji}^{new} = x_i$$

لذا، تصبح مصفوفة وزن أعلى لأسفل:

$$\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{0} \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

3. في حالة شعاع الدخل الثانـــي (0, 0, 0, 1)، كرر الخطوات من 4-13

4. ضع تفعيلات كل عقد الطبقة F2 بقيمة الصفر، ضع تفعيلات كل عقد الطبقة F1 بقيمة

شعاع الدخل $\mathbf{I} = (0, 0, 0, 1)$

5. احسب نظيم شعاع الدخل:

$$\|\mathbf{I}\| = 1$$

6. حساب تفعيلات كل عقدة في طبقة الدخل F1:

$$\mathbf{X} = (0,0,0,1)$$

7. حساب دخل كل عقدة في الطبقة F2:

$$y_1 = 0.67(0) + 0.67(0) + 0.67(0) + 0(1) = 0.0$$

$$y_2 = 0.2(0) + 0.2(0) + 0.2(0) + 0.2(1) = 0.2$$

$$y_3 = 0.2(0) + 0.2(0) + 0.2(0) + 0.2(1) = 0.2$$

8. مادام التصفير صحيحاً (true)، كرر الخطوات 9-12

9. لما كان للوحدتين $Y_2$ و$Y_3$ الدخل نفسه فإن $J = 2$.

10. إعادة حساب تفعيلات الطبقة F1:

الآن $T_2 = (1, 1, 1, 1)$ ومن ثم $\mathbf{X} = (0, 0, 0, 1)$

11. حساب نظيم الشعاع **X**:

$$\|\mathbf{X}\| = 1$$

12. اختبار شرط التصفير:

$$|\mathbf{X}|/|\mathbf{I}| = 1 > 0.4$$

ويكون التصفير خطأ. نفذ الخطوة 13

13. تحديث $\mathbf{v}_2$: في حالة $L = 2$، تكون أوزان التوازن:

$$\upsilon_{i2}^{new} = 2x_i/(1 + \|\mathbf{X}\|)$$

لذا، تصبح مصفوفة وزن أدنـــى لأعلى:

$$\begin{bmatrix} 0.67 & 0 & 0.2 \\ 0.67 & 0 & 0.2 \\ 0 & 0 & 0.2 \\ 0 & 1 & 0.2 \end{bmatrix}$$

تحديث $\mathbf{w}_2$: قيم وزن التعليم السريع

$$w_{2i}^{new} = x_i$$

لذا، تصبح مصفوفة وزن أعلى- لأسفل:

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

3. في حالة شعاع الدخل الثالث (1, 0, 0, 0)، كرر الخطوات من 4-13
4. ضع تفعيلات كل عقد الطبقة F2 بقيمة الصفر، ضع تفعيلات كل عقد الطبقة F1 بقيمة شعاع الدخل $\mathbf{I} = (1, 0, 0, 0)$
5. احسب نظيم شعاع الدخل:

$$\|\mathbf{I}\| = 1$$

6. حساب تفعيلات كل عقدة في طبقة الدخل F1:

$$\mathbf{X} = (1, 0, 0, 0)$$

7. حساب دخل كل عقدة في الطبقة F2:

$$y_1 = 0.67(1) + 0.67(0) + 0(0) + 0(0) = 0.67$$
$$y_2 = 0(1) + 0(0) + 0(0) + 1(0) = 0.0$$
$$y_3 = 0.2(1) + 0.2(0) + 0.2(0) + 0.2(0) = 0.2$$

8. مادام التصفير صحيحاً (true)، كرر الخطوات 9-12
9. لما كان للوحدات $Y_1$ الدخل الأعظم، فإن $J = 1$.
10. إعادة حساب تفعيلات الطبقة F1:

الآن $T_1 = (1, 1, 0, 0)$ ومن ثم $\mathbf{X} = (1, 0, 0, 0)$

11. حساب نظيم الشعاع x:

$$\|\mathbf{X}\| = 1$$

12. اختبار شرط التصفير:

$$|\mathbf{X}|/|\mathbf{I}| = 1 > 0.4$$

ويكون التصفير خطأ. نفذ الخطوة 13

13. تحديث $\mathbf{v}_1$: في حالة L = 2 تكون أوزان التوازن:

$$v_{i1}^{new} = 2x_i /(1+\|\mathbf{X}\|)$$

لذا، تصبح مصفوفة وزن أسفل لأعلى:

$$\begin{bmatrix} 1 & 0 & 0.2 \\ 0 & 0 & 0.2 \\ 0 & 0 & 0.2 \\ 0 & 1 & 0.2 \end{bmatrix}$$

تحديث $\mathbf{w}_1$: قيم وزن التعليم السريع

$$w_{1i}^{new} = x_i$$

لذا، تصبح مصفوفة وزن أعلى لأسفل:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

3. في حالة شعاع الدخل الرابع (0, 0, 1, 1)، كرر الخطوات من 4-13
4. ضع تفعيلات كل العقد في الطبقة F2 بقيمة الصفر، وضع تفعيلات كل العقد في الطبقة F1 بقيمة شعاع الدخل **I** = (0, 0, 1, 1)
5. احسب نظيم شعاع الدخل:

$$\|\mathbf{I}\| = 2$$

6. حساب تفعيلات كل عقدة في طبقة الدخل F1:

$$\mathbf{X} = (0, 0, 1, 1)$$

7. حساب دخل كل عقدة في الطبقة F2:

$$y_1 = 1(0) + 0(0) + 0(1) + 0(1) = 0.0$$

$$y_2 = 0(0) + 0(0) + 0(1) + 1(1) = 1.0$$

$$y_3 = 0.2(0) + 0.2(0) + 0.2(1) + 0.2(1) = 0.4$$

8. مادام التصفير صحيحاً (true)، كرر الخطوات 9-12
9. لما كان للوحدات $Y_2$ الدخل الأكبر فإن $J = 2$.
10. إعادة حساب تفعيلات الطبقة F1:

الآن $T_2 = (0, 0, 0, 1)$ ومن ثم $\mathbf{X} = (0, 0, 0, 1)$

11. حساب نظيم الشعاع $\mathbf{x}$:

$$\|\mathbf{X}\| = 1$$

12. اختبار شرط التصفير:

$$|\mathbf{X}|/|\mathbf{I}| = 0.5 > 0.4$$

ويكون التصفير خطأ. نفذ الخطوة 13

13. تحديث $\mathbf{v}_2$: لن يكون هناك تغير في مصفوفة أوزان أدنـــى لأعلى:

$$v_{i2}^{new} = 2x_i /(1+\|\mathbf{X}\|)$$

لذا ، تصبح مصفوفة وزن أدنـــى لأعلى:

$$\begin{bmatrix} 1 & \mathbf{0} & 0.2 \\ 0 & \mathbf{0} & 0.2 \\ 0 & \mathbf{0} & 0.2 \\ 0 & \mathbf{1} & 0.2 \end{bmatrix}$$

ولن يكون هناك تغير في مصفوفة أعلى لأدنى. لذا، تصبح مصفوفة وزن أعلى لأدنـــى:

$$\begin{bmatrix} \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

13. اختبار شرط التوقف،( هذه نهاية أول دور تدريب).

يمكن للقارئ أن يتأكد أنه لن يحدث تعليم أبعد من ذلك على التمثيلات المتتالية لهذه الأشعة، بصرف النظر عن الدرجة التـــي تمثل فيها. بالاعتماد على درجة تمثيل النماذج، يمكن أن يلزم أكثر من دور واحد، لكن مصفوفات الأوزان تستقر بسرعة جداً.

**مثال 2:**

الآن سنحاول تنفيذ المثال السابق في حالة وسيط احتراس متوسط، لذا سنقدم إلى الشبكة نفس أشعة الدخل وبنفس الترتيب. وسنعتمد وسيط الاحتراس بقيمة 0.7 وسيجري تدريب الأشعة (1, 1, 0, 0) و(0, 0, 0, 1) و(1, 0, 0, 0) كما في المثال السابق، وستكون مصفوفة وزن أدنـــى لأعلى:

$$\begin{bmatrix} 1 & 0 & 0.2 \\ 0 & 0 & 0.2 \\ 0 & 0 & 0.2 \\ 0 & 1 & 0.2 \end{bmatrix}$$

وستكون مصفوفة وزن أعلى لأدنى:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

وستكون النتيجة مختلفة في حالة شعاع الدخل الرابع $\mathbf{I} = (0, 0, 1, 1)$ على النحو التالي:

3. في حالة شعاع الدخل الرابع (0, 0, 1, 1)، كرر الخطوات من 4 إلى 13.

4. ضع تفعيلات كل العقد في الطبقة F2 بقيمة الصفر، وضع تفعيلات كل العقد في الطبقة F1 بقيمة شعاع الدخل $\mathbf{I} = (0, 0, 1, 1)$

5. احسب نظيم شعاع الدخل:

$$\|\mathbf{I}\| = 2$$

6. حساب تفعيلات كل عقدة في طبقة الدخل F1:

$$\mathbf{X} = (0, 0, 1, 1)$$

7. حساب دخل كل عقدة في الطبقة F2:

$$y_1 = 1(0) + 0(0) + 0(1) + 0(1) = 0.0$$

$$y_2 = 0(0) + 0(0) + 0(1) + 1(1) = 1.0$$

$$y_3 = 0.2(0) + 0.2(0) + 0.2(1) + 0.2(1) = 0.4$$

8. مادام التصفير صحيحاً (true)، كرر الخطوات 9-12

9. لما كان للوحدة $Y_2$ الدخل الأكبر فإن $J = 2$.

10. إعادة حساب تفعيلات الطبقة F1:

الآن $T_2 = (0, 0, 0, 1)$ ومن ثم $\mathbf{X} = (0, 0, 0, 1)$

11. حساب نظيم الشعاع **X**:

$$\|\mathbf{X}\| = 1$$

12. اختبار شرط التصفير:

$$|\mathbf{X}|/|\mathbf{I}| = 0.5 < 0.7$$

التصفير صحيح، و $Y_2$ مخمدة:

$$y_2 = -1.0$$

نفذ الخطوة 8.

8. مادام التصفير صحيحاً (true)، كرر الخطوات 9-12

9. قيم الطبقة F2 ستكون:

$$y_1 = 0.0 \quad y_2 = -1.0 \quad y_3 = 0.4$$

لما كان للوحدة $Y_3$ الدخل الأكبر فإن $J = 3$.

10. إعادة حساب تفعيلات الطبقة F1:

الآن $T_3 = (1, 1, 1, 1)$ ومن ثم $\mathbf{X} = (0, 0, 1, 1)$

11. حساب نظيم الشعاع X:

$$\|\mathbf{X}\| = 2$$

12. اختبار شرط التصفير:

$$|\mathbf{X}|/|\mathbf{I}| = 1.0 > 0.7$$

التصفير خطأ، ومن ثم نفذ الخطوة 13.

13. تحديث $\mathbf{v}_3$: تصبح مصفوفة أوزان أدنـــى لأعلى:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0.6 \\ 0 & 1 & 0.6 \end{bmatrix}$$

تحديث $\mathbf{w}_3$: تصبح مصفوفة وزن أعلى لأدنى:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

14. اختبار شرط التوقف.

الآن، عندما يقدم الشعاع الأول يقدم ثانيةً، لن يتحقق معيار الاحتراس لأي عقـــدة في F2. ويمكن أن يقرر المصمم إضافة عقدة F2 جديدة، أو تصنيف شعاع الدخل الأول على أنه أعزل، أو استعمال وسيط احتراس أخفض. وخلافاً لبعض الشبكات العصبونية الأخرى، لن تجبر شبكة نظرية الطنين المتكيف الأولي آلياً كل أشعة الدخل على أن تكون ضمن تجمعات إذا لم تكن متشابهة بقدر كافٍ.

**مثال 3:**

الآن سنناقش استعمال شبكة نظرية الطنين المتكيف الأولى لتجميع النماذج الموضحة في (الشكل 3.14) باستعمال اختيار تمثيلي لقيم مختلفة من وسيط الاحتراس، وتراتيب دخل

مختلفة للنماذج، وقيم مختلفة للعدد الأعظمي لوحدات التجمع ( وحدات الطبقة F2). لاحظ أن شعاع الوزن لكل تجمع يعكس كل النماذج المتوضعة على التجمع خلال التدريب، وأن الشبكة لن تنسى النماذج التـــي توضعت على الوحدة، ثم تحركت إلى وحدة أخرى. الأوزان النهائية المرافقة لكل تجمع ستكون مصفوفة ثنائية البعد، باعتبار أن نماذج الدخل ممثلة بنماذج ثنائية البعد.

في هذا المثال سنحاول تجميع الأحرف من ثلاث تشكيلات مختلفة مع وسيط احتراس منخفض.

**باستعمال ترتيب النماذج على النحو التالي:**

$A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3, D_1, D_2, D_3, E_1, E_2, E_3, J_1, J_2, J_3, K_1, K_2, K_3$

وقيمة وسيط الاحتراس تساوي 0.3، وعدد أعظمي من وحدات التجمع يساوي 10، حصلنا على النتائج في تشكيلة التجمع المستقر (لاتغير للأوزان) بعد ثلاثة أدوار تدريب.

كان توضع النماذج خلال التدريب على النحو التالي:

| التجمع | الدور الأول | الدور الثانـــي | الدور الثالث |
|---|---|---|---|
| 1 | $A_1, A_2, A_3$ | $A_1, A_2, A_3$ | $A_1, A_2, A_3$ |
| 2 | $B_1, B_2, B_3$<br>$C_1, C_2, C_3$<br>$J_1$ | | |
| 3 | $D_1, D_2, D_3$<br>$E_1, E_2, E_3$ | $B_1, B_2, B_3$<br>$C_1, C_2, C_3$ | $C_1, C_2, C_3$ |
| 4 | $J_2, J_3$ | $J_1, J_2, J_3$ | $J_1, J_2, J_3$ |
| 5 | $K_1, K_2$ | $K_1, K_2$ | $K_1, K_2$ |
| 6 | $J_3$ | $D_1, D_2, D_3$<br>$K_3$ | $D_1, D_2, D$<br>$K_3$ |
| 7 | | $E_1, E_2, E_3$ | $B_1, B_2, B_3$<br>$E_1, E_2, E_3$ |

الدخل من التشكيلة الأولى:

```
00##000   ######0   00#####   #####00   #######   000####   ###00##
000#000   0#0000#   0#0000#   0#000#0   0#0000#   00000#0   0#00#00
000#000   0#0000#   #000000   0#0000#   0#00000   00000#0   0#0#000
00#0#00   0#0000#   #000000   0#0000#   0#0#000   00000#0   0##0000
00#0#00   0#####0   #000000   0#0000#   0###000   00000#0   0##0000
0#####0   0#0000#   #000000   0#0000#   0#0#000   00000#0   0#0#000
0#000#0   0#0000#   #000000   0#0000#   0#00000   0#000#0   0#00#00
0#000#0   0#0000#   0#0000#   0#000#0   0#0000#   0#000#0   0#000#0
###0###   ######0   00####0   #####00   #######   00###00 . ###00##
A......   .B.....   ..C....   ...D...   ....E..   ......J.  ......K
```

الدخل من التشكيلة الثانية:

```
000#000   ######0   00###00   #####00   #######   00000#0   #0000#0
000#000   #00000#   0#000#0   #0000#0   #000000   00000#0   #000#00
000#000   #00000#   #00000#   #00000#   #000000   00000#0   #00#000
00#0#00   #00000#   #000000   #00000#   #000000   00000#0   #0#0000
00#0#00   ######0   #000000   #00000#   #####00   00000#0   ##00000
0#000#0   #00000#   #000000   #00000#   #000000   00000#0   #0#0000
0#####0   #00000#   #00000#   #00000#   #000000   0#000#0   #00#000
0#000#0   #00000#   0#000#0   #0000#0   #000000   0#000#0   #000#00
0#000#0   ######0   00###00   #####00   #######   00###00   #0000#0
A......   .B.....   ..C....   ...D...   ....E..   ......J.  ......K
```

الدخل من التشكيلة الثالثة:

```
000#000   ######0   00###0#   #####00   #######   0000###   ###00##
000#000   0#0000#   0#000##   0#000#0   0#0000#   00000#0   0#000#0
00#0#00   0#0000#   #00000#   0#0000#   0#00#00   00000#0   0#00#00
00#0#00   0#0000#   #000000   0#0000#   0####00   00000#0   0#0#000
0#000#0   0#####0   #000000   0#0000#   0#00#00   00000#0   0##0000
0#####0   0#0000#   #000000   0#0000#   0#00000   00000#0   0#0#000
#00000#   0#0000#   #00000#   0#0000#   0#00000   00000#0   0#00#00
#00000#   0#0000#   0#000#0   0#000#0   0#0000#   0#000#0   0#000#0
##000##   ######0   00###00   #####00   #######   00###00   ###00##
A......   .B.....   ..C....   ...D...   ....E..   .....J.   ......K
```

**الشكل 3.14** نماذج دخل التدريب

الأوزان النهائية:

| التجمع 1 | التجمع 2 | التجمع 3 | التجمع 4 | التجمع 5 | التجمع 6 | التجمع 7 |
|---|---|---|---|---|---|---|
| 000#000 | 000##00 | 00###00 | 00000#0 | #0000#0 | ###0000 | ######0 |
| 000#000 | 0000000 | 0000000 | 00000#0 | 0000#00 | 00000#0 | 0000000 |
| 0000000 | 0000000 | 0000000 | 00000#0 | 000#000 | 0000000 | 0000000 |
| 00#0#00 | 0000000 | 0000000 | 00000#0 | 00#0000 | 0000000 | 0000000 |
| 0000000 | 0000000 | 0000000 | 00000#0 | 0#00000 | 0000000 | 0#00000 |
| 0#000#0 | 0000000 | 0000000 | 00000#0 | 0000000 | 0000000 | 0000000 |
| 0000000 | 0000000 | 0000000 | 00000#0 | 0000000 | 0000000 | 0000000 |
| 0000000 | 0000000 | 0000000 | 0#000#0 | 0000000 | 00000#0 | 0000000 |
| #00000# | 00###00 | 00###00 | 00###00 | #0000#0 | ###0000 | ######0 |

**باستعمال ترتيب دخل النماذج التالي:**

$A_1, B_1, C_1, D_1, E_1, J_1, K_1, A_2, B_2, C_2, D_2, E_2, J_2, K_2, A_3, B_3, C_3, D_3, E_3, J_3, K_3$

وباستعمال وسيط احتراس يساوي 0.3 مع 10 عقد تجمع متوفرة، حصلنا على النتائج في تشكيل التجمع المستقر (لا تغير للأوزان) بعد دوري تدريب.

كان توضع النماذج خلال التدريب على النحو التالي:

| التجمع | الدور الأول | الدور الثاني |
|---|---|---|
| 1 | $A_1, B_1, C$ | $C_1$ |
| 2 | $D_1, E_1, J_1$ | $J_2$ |
| | $C_2, J_2$ | |
| 3 | $K_1, A_2,$ | $A_1, A_2$ |
| 4 | $B_2, D_2, E_2, K_2$ | $B_2, D_2, E_2, K_2$ |
| 5 | $A_3, B_3, E_3$ | $A_3$ |
| 6 | $C_3, D_3, J_3$ | $J_1, C_2, C_3, J_3$ |
| 7 | $K_3$ | $B_1, D_1, E_1, K_1$ |
| | | $B_3, D_3, E_3, K_3$ |

**مثال 4:**

باستعمال وسيط احتراس يساوي 0.7، لكن ما يزال مسموحاً استخدام عدد أعظمي يساوي 10 وحدات تجمع، حصلنا على النتائج في تشكيل المقاطع المستقر (لا تغير في الأوزان) بعد دورين فقط من التدريب. على أية حال، بعض النماذج لا يمكن أن توضع على التجمعات (سيظهر في النتائج عبارة عدم استطاعة التجمع CNC؛ (Could Not Cluster).

الأوزان النهائية :

| التجمع 1 | التجمع 2 | التجمع 3 | التجمع 4 | التجمع 5 | التجمع 6 | التجمع 7 |
|---|---|---|---|---|---|---|
| 00##000 | 0000000 | 0000000 | #000000 | 000#000 | 0000#00 | ###0000 |
| 0000000 | 0000000 | 0000000 | #000000 | 0000000 | 00000#0 | 0#00000 |
| 0000000 | 0000000 | 000#000 | #000000 | 0000000 | 0000000 | 0#00000 |
| 0000000 | 0000000 | 00#0000 | #000000 | 00#0#00 | 0000000 | 0#00000 |
| 0000000 | 0000000 | 00#0000 | #000000 | 0#00000 | 0000000 | 0#00000 |
| 0000000 | 0000000 | 0#00000 | #000000 | 0#00000 | 0000000 | 0#00000 |
| 0000000 | 0000000 | 0#00000 | #000000 | 0000000 | 0000000 | 0#00000 |
| 0#00000 | 0#00000 | 0000000 | #000000 | 0000000 | 0#000#0 | 0#00000 |
| 00#0##0 | 00###00 | #00000# | #000000 | ##000#0 | 00###00 | ###0000 |

**ترتيب دخل النماذج لهذا المثال كان على النحو التالي:**

$A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3, D_1, D_2, D_3, E_1, E_2, E_3, J_1, J_2, J_3, K_1, K_2, K_3$

| التجمع | الدور الأول | الدور الثاني |
|---|---|---|
| 1 | $A_1,A_2$ | $A_2$ |
| 2 | $A_3$ | $A_3$ |
| 3 | $B_1,B_2$ | $B_1,B_2$ |
| 4 | $B_3,D_1,D_3$ | $B_3,D_1,D_3$ |
| 5 | $C_1,C_2,K_2$ | $C_1,C_2$ |
| 6 | $C_3$ | $C_3$ |
| 7 | $D_2$ | $D_2$ |
| 8 | $E_1,E_3,K_1$ $K_3$ | $E_1,E_3$ |
| 9 | $E_2$ | $E_2$ |
| 10 | $J_1,J_2,J_3$ | $J_1,J_2,J_3$ |
| CNC | $K_2$ | $A_1,E_1,E_3,K_2$ |

**مثال 5:**

باستعمال نفس وسيط الاحتراس كما في المثال 4 بقيمة تساوي 0.7، لكن بأخذ عدد وحدات تجمع 15 وحدة كعدد أعظمي، نتائج عملية تشكيل التجمعات المستقرة (لا تغير في الأوزان) أعطيت بعد دورين اثنين للتدريب. التجمعات المتشكلة كانت أقل بالاعتماد على ترتيب دخل النماذج في حالة وسيط الاحتراس العالي مقارنة مع النتائج في المثال 3 في حالة وسيط احتراس منخفض.

الأوزان النهائية:

| التجمع 1 | التجمع 2 | التجمع 3 | التجمع 4 | التجمع 5 |
|---|---|---|---|---|
| 000#000 | 000#000 | ######0 | #####00 | 00###00 |
| 000#000 | 000#000 | 000000# | 0#00000 | 0#00000 |
| 000#000 | 00#0#00 | 000000# | 0#0000# | #000000 |
| 00#0#00 | 00#0#00 | 000000# | 0#00000 | #000000 |
| 00#0#00 | 0#000#0 | 0#####0 | 0#0000# | #000000 |
| 0#000#0 | 0#####0 | 000000# | 0#0000# | #000000 |
| 0#000#0 | #00000# | 000000# | 0#0000# | #000000 |
| 0000000 | #00000# | 000000# | #000000 | 0#00000 |
| #00000# | ##000## | ######0 | #####00 | 00###00 |

| التجمع 6 | التجمع 7 | التجمع 8 | التجمع 9 | التجمع10 |
|---|---|---|---|---|
| 00###0# | #####00 | ###00## | ####### | 00000#0 |
| 0#000## | #0000#0 | 0#00000 | #000000 | 00000#0 |
| #00000# | #00000# | 0#00000 | #000000 | 00000#0 |
| #000000 | #00000# | 0#00000 | #0#0000 | 00000#0 |
| #000000 | #00000# | 0#00000 | #####00 | 00000#0 |
| #000000 | #00000# | 0#00000 | #000000 | 00000#0 |
| #00000# | #00000# | 0#00000 | #000000 | 00000#0 |
| 0#000#0 | #0000#0 | 0#00000 | #000000 | 0#000#0 |
| 00###00 | #####00 | ###00## | ####### | 00###00 |

**باستعمال أول ترتيب دخل للنماذج:**

$A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3, D_1, D_2, D_3, E_1, E_2, E_3, J_1, J_2, J_3, K_1, K_2, K_3$

حصلنا على النتائج التالية:

| التجمع | الدور الأول | الدور الثانـــي |
|---|---|---|
| 1 | $A_1, A_2$ | $A_2$ |
| 2 | $A_3$ | $A_3$ |
| 3 | $B_1, B_2$ | $B_1, B_2$ |
| 4 | $B_3, D_1, D_3$ | $B_3, D_1, D_3$ |
| 5 | $C_1, C_2$ | $C_1, C_2$ |
| 6 | $C_3$ | $C_3$ |
| 7 | $D_2$ | $D_2$ |
| 8 | $E_1, E_3, K_1, K_3$ | $K_1, K_3$ |
| 9 | $E_2$ | $E_2$ |
| 10 | $J_1, J_2, J_3$ | $J_1, J_2, J_3$ |
| 11 | $K_2$ | $K_2$ |
| 12 | | $A_1$ |
| 13 | | $E_1, E_3$ |

الأوزان النهائية:

| التجمع 1 | التجمع 2 | التجمع 3 | التجمع 4 | التجمع 5 |
|---|---|---|---|---|
| 000#000 | 000#000 | ######0 | #####00 | 00###00 |
| 000#000 | 000#000 | 000000# | 0#00000 | 0#00000 |
| 000#000 | 00#0#00 | 000000# | 0#0000 | #000000 |
| 00#0#00 | 00#0#00 | 000000# | 0#00000 | #000000 |
| 00#0#00 | 0#000#0 | 0#####0 | 0#0000# | #000000 |
| 0#000#0 | 0#####0 | 000000# | 0#0000# | #000000 |
| 0#000#0 | #00000# | 000000# | 0#0000# | #000000 |
| 0000000 | #00000# | 000000# | #000000 | 0#00000 |
| #00000# | ##000## | ######0 | #####00 | 00###00 |

| التجمع 6 | التجمع 7 | التجمع 8 | التجمع 9 | التجمع10 |
|---|---|---|---|---|
| 00###0# | #####00 | ###00## | ####### | 00000#0 |
| 0#000## | #0000#0 | 0#00000 | #000000 | 00000#0 |
| #00000# | #00000# | 0#00000 | #000000 | 00000#0 |
| #000000 | #00000# | 0#00000 | #0#0000 | 00000#0 |
| #000000 | #00000# | 0#00000 | #####00 | 00000#0 |
| #000000 | #00000# | 0#00000 | #000000 | 00000#0 |
| #00000# | #00000# | 0#00000 | #000000 | 00000#0 |
| 0#00##0 | #0000#0 | 0#00000 | #000000 | 0#000#0 |
| 00###00 | #####00 | ###00## | ####### | 00###00 |

| التجمع 11 | التجمع 12 | التجمع 13 |
|---|---|---|
| #0000#0 | 00##000 | ####### |
| #000#00 | 000#000 | 0#0000# |
| #00#000 | 000#000 | 0#00000 |
| #0#0000 | 00#0#00 | 0#0#000 |
| ##00000 | 00#0#00 | 0#00000 |
| #0#0000 | 0#####0 | 0#00000 |
| #00#000 | 0#000#0 | 0#00000 |
| #000#00 | 0#000#0 | 0#0000# |
| #0000#0 | ###0### | ######## |

**النتائج من أجل ترتيب دخل النماذج الثانـــي التالي:**

$A_1$ ,$B_1$ ,$C_1$ ,$D_1$ ,$E_1$ ,J1 ,$K_1$,$A_2$ ,$B_2$ ,$C_2$ ,$D_2$ ,$E_2$ ,$J_2$ ,$K_2$, $A_3$ ,$B_3$ ,$C_3$ ,$D_3$ ,$E_3$ ,$J_3$ ,$K_3$

ستكون مشابهة تماماً (ولكن ليست متطابقة) للنتائج في حالة ترتيب دخل النماذج الأول:

| التجمع | الدور الأول | الدور الثانـــي |
|---|---|---|
| 1 | $A_1, A_2$ | $A_2$ |
| 2 | $B_1, D_1, D_3$ | $B_1, D_1, D_3$ |
| 3 | $C_1, C_2$ | $C_1, C_2$ |
| 4 | $E_1, K_1, K_3$ | $E_1, K_1, K_3$ |
| 5 | $J_1, J_2, J_3$ | $J_1, J_2, J_3$ |
| 6 | $B_2, D_2$ | $B_2, D_2$ |
| 7 | $E_2$ | $E_2$ |
| 8 | $K_2$ | $K_2$ |
| 9 | $A_3$ | $A_3$ |
| 10 | $B_3, E_3$ | $B_3, E$ |
| 11 | $C_3$ | $C_3$ |
| 12 | | $A_1$ |

```
000#000
000#000
000#000
00#0#00
00#0#00
0#000#0
0#000#0
0000000
#00000#
```
التجمع 1

```
#####00
0#00000
0#0000#
0#0000#
0##0000
0#0000#
0#0000#
0#00000
#####00
```
التجمع 2

```
00###00
0#00000
#000000
#000000
#000000
#000000
#000000
0#00000
00###00
```
التجمع 3

```
###00##
0#00000
0#00000
0#00000
0##0000
0#0#000
0#00000
0#00000
###00##
```
التجمع 4

```
00000#0
00000#0
00000#0
00000#0
00000#0
00000#0
00000#0
0#000#0
00###00
```
التجمع5

```
#####00
#000000
#00000#
#00000#
#000000
#00000#
#00000#
#000000
#####00
```
التجمع 6

```
#######
#000000
#000000
#0#0000
#####00
#000000
#000000
#000000
#######
```
التجمع 7

```
#0000#0
#000#00
#00#000
#0#0000
##00000
#0#0000
#00#000
#000#00
#0000#0
```
التجمع8

```
000#000
000#000
00#0#00
00#0#00
0#000#0
0#####0
#00000#
#00000#
##000##
```
التجمع 9

```
######0
0#0000#
0#00000
0####00
0#00000
0#00000
0#00000
0#0000#
######0
```
التجمع 10

```
00###0#
0#000##
#00000#
#000000
#000000
#000000
#00000#
0#000#0
00###00
```
التجمع11

```
00##000
000#000
000#000
00#0#00
00#0#00
0#####0
0#000#0
0#000#0
###0###
```
التجمع 12

## 5.14 بنية شبكة نظرية الطنين المتكيف الثانية
## The ART2 network architecture

تعتبر شبكات نظرية الطنين المتكيف الثانية تعميماً لشبكات نظرية الطنين المتكيف الأولى المدروسة في الفقرات السابقة. هذه الشبكات قادرة على تعلم وتنظيم نماذج الدخل بقيم ثنائية وبقيم حقيقية (وهذه أول خاصية تميزها عن شبكات ART1).

نظام توصيل شبكات نظرية الطنين الثانية مشابه للشبكات الأولى باستثناء أن عقد الدخل تكون أكثر تعقيداً. تتألف عقد طبقة الدخل F1 في شبكات نظرية الطنين المتكيف الثانية، فعلياً، من شبكة جزئية صغيرة فيها ست عقد (Q, P, S, U, X, T) تعمل كمنطقة حرة (Buffer) بين إشارة الدخل وإشارة التوقع أعلى لأدنى. تنفذ هذه العقد أيضاً وظائف المعايرة والانسجام على إشارات الدخل وإشارات التغذية العكسية من الطبقة F2. إن طبقة الفئة أو طبقة التجمع F2 في هذه الشبكات هي نفسها في شبكات النوع الأول المدروسة في الفقرات السابقة.

## 6.14 ديناميكيات شبكة نظرية الطنين المتكيف الثانية ART Dynamics

يوضح (الشكل 4.14) بنية مبسطة جداً لهذه الشبكة في حالة عقدة واحدة فقط j (من بين m عقدة لها نفس نموذج التوصيل) في الطبقة F2، حيث وسع الدخل الوحيد فقط (من بين n دخلاً له نفس نموذج التوصيل الموضح في الشكل) لعقدة الطبقة F1 لتنفيذ وظيفة عقد الشبكة الجزئية المحتواة ضمن العقدة رقم i في الطبقة F1. الوصلات بين العقد المختلفة التـــي لم يظهر عليها وزن في الشكل تعتبر تحويلات بدون أي تعديل، كما أن الرمز $\xrightarrow{o}$ يشير إلى المعايرة؛ أي جعل شعاع تفعيلات الوحدتين الموصل بينهما بهذا الرمز معيارياً بطول يســـاوي الواحد. الوصلات بين العقدتين T إلى U وبين العقدتين Q إلى V لهما أوزان ثابتة a و b على الترتيب.

تمر إشارة الدخل $I_i$، $i = 1,2,\cdots,n$، أدنـــى لأعلى إلى العقدة رقم j في الطبقة F2، $j = 1,2,\cdots,m$، يعطى بـــ:

$$I_i \to t_i \to x_i \to s_i \to u_i \to p_i \to S_j$$

حيث $$S_j = \sum_{i=1}^{n} p_i v_{ij}$$

هو مجموع كل المداخل المثقلة إلى العقدة j من الطبقة F1.

ممر إشارة أعلى لأدنـــى من الخرج $y_j$ للعقدة رقم j في الطبقة F2 إلى العقدة رقم i في الطبقة F1 هو التالي:

$$y_j \to p_i \to q_i \to s_i \to u_j$$

لاحظ أن الإشارة المستقبلة عند $u_i$ ، ستتغير بوجه عام نتيجة لعودة إشارة الممر. لوحظ أيضاً، أن وظيفة التحكم بربح الانتباه AGC الموصوفة في شبكات نظرية الطنين المتكيف الأولى وزعت عبر عقد عديدة في شبكة الطنين المتكيف الثانية. نفّذ التحكم في ربح الانتباه بناءً على وصلات مارة عبر العقد الغامقة الصغيرة في (الشكل 4.14)؛ نوى التحكم في الربح التـــي تخمد عقدها المنشودة بالتناسب مع معيار L2 لمخارج عقد منبع التفعيل.

تكتب فعالية خرج العقدة i، $i = 1,2,\cdots,n$ ، $o_i$ في الطبقة F1 في نفس الشكل العام المكتوب في حالة شبكة الطنين المتكيف الأولى، أي:

$$\varepsilon \frac{do_i}{dt} = -Ao_i + (1 - Bo_i)J_i^+ - (C + Do_i)J_i^- \qquad (15.14)$$

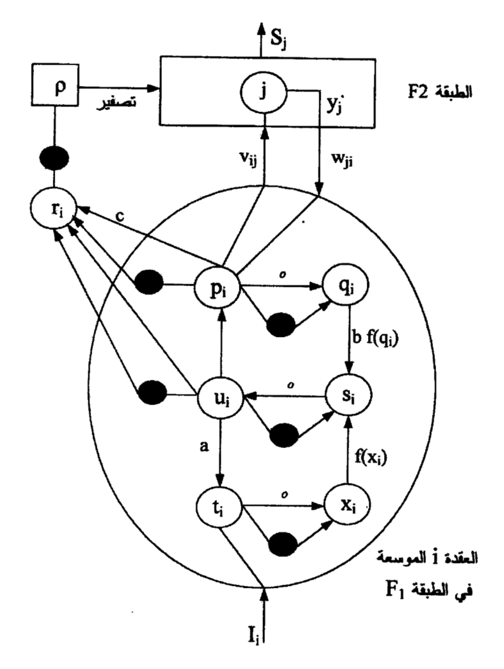
حيث، وكما في حالة شبكة الطنين المتكيف الأولى، $J_i^+$ هو دخل التهييج الكلي، و $J_i^-$ هو دخل التخميد الكلي للعقدة I، و$\varepsilon$ ثابت صغير موجب يعبر عن نسبة أزمنة تراخي ذاكرة الأجل القصير وذاكرة الأجل الطويل.

في حالة شبكة نظرية الطنين المتكيف الثانية (بالمطابقة مع (الشكل 4.14)) يمكن تبسيط المعادلة (15.14) بوضع $B = C = 0$ وافتراض $\varepsilon \to 0$.

الشكل الفريد للمعادلة (15.14) يمكن أن يكتب من جديد بالشكل:

$$o_i = \frac{J_i^+}{A + DJ_i^-} \qquad (16.14)$$

باستعمال هذا الشكل، يمكن إعطاء المعادلات الضابطة لديناميكيات العقدة $i$ في الطبقة F1 (بشبكة جزئية موسعة كما في (الشكل 4.14)).




**الشكل 4.14**: شبكة نظرية الطنين المتكيف الثانية مع عقدة **i** مفردة موسعة في الطبقة **F1**

وستكون في حالة تدفق الإشارة من أدنـــى لأعلى كما يلي:

$$t_i = I_i + au_i \quad (17.14)$$

$$x_i = t_i/(e + \|\mathbf{t}\|) \quad (18.14)$$

$$s_i = f(x_i) + bf(q_i) \quad (19.14)$$

$$u_i = s_i/(e + \|\mathbf{s}\|) \quad (20.14)$$

$$q_i = p_i/(e + \|\mathbf{p}\|) \quad (21.14)$$

$$p_i = u_i + \sum_j g(y_j) w_{ji} \quad (22.14)$$

الثابت a هو الوزن بين العقدتين U وT، $0 < a < 1$، والثابت b هو الوزن بين العقدتين Q وS، $b > 0$، والثابت e وسيط بقيمة صغيرة يستعمل كإجراء وقائي لمنع التقسيم على الصفر عندما يكون نظيم الشعاع يساوي الصفر، $e > 0$، و$\|x\|$ نظيم الشعاع x، وf(x) تابع تفاضلي مستمر يعطى بالمعادلة التالية:

$$f(x) = \frac{2\theta x^2}{x^2 + \theta^2} \quad \text{في حالة } 0 \le x \le \theta$$

$$f(x) = x \quad \text{ماعدا ذلك} \quad (23.14)$$

باعتبار قيم $x_i$ و $q_i$ بين الصفر والواحد، وقيمة f تكون أيضاً محددة بالصفر والواحد (عملياً يمكن أن تبسط المعادلات (17.14) حتى (22.14) بوضع e بقيمة الصفر، و $\|x\| = \|q\| = \|u\| = 1$). التابع $g(y_j)$ في المعادلة (22.14) هو تابع الاختيار التنافسي الذي يجعل الخيار فيما بين عقد الطبقة F2، بالاعتماد على الشدات النسبية للإشارات $S_j$ عند الطبقة F2، هذا يعني أنه:

إذا كان {العقدة رقم k لم تصفر: $S_j = \max\{S_k$ فإن $g(y_j) = d$

$g(y_j) = 0$ ماعدا ذلك (24.14)

حيث $y_j$ فعالية العقدة رقم j في الطبقة F2، والثابت d هو تفعيل وحدة الطبقة F2 الرابحة، $0 < d < 1$، قيمته النموذجية تساوي 0.9، ويجب أن يختار كل من c و d لتحقيق المتراجحة التالية:

$$[cd/(1-d)] \le 1$$

لمنع حدوث التصفير خلال تجربة التعليم، والنسبة ستكون مختارة قريبة من الواحد لإعطاء مجال فعال أكبر لوسيط الاحتراس.

من المعادلة (24.14) نستطيع تبسيط المعادلة (22.14) كما يلي:

إذا كانت F2 غير فعالة فإن $p_i = u_i$

(25.14) إذا كانت العقدة J في الطبقة F2 فعالة فإن $p_i = u_i + d w_{Ji}$

تنفذ عملية الانسجام والتصفير في شبكات نظرية الطنين المتكيف الثانية بواسطة نظام

التوجيه الجزئي مع الشعاع $\mathbf{r} = (r_1, r_2, \cdots, r_n)$ ، حيث

$$r_i = \frac{(u_i + c\,p_i)}{(\|\mathbf{u}\| + \|\mathbf{cp}\|)} \qquad (c > 0) \qquad (26.14)$$

حيث الثابت c هو الوزن المستعمل في اختبار التصفير، قيمته النموذجية تساوي 0.1.

تعطي قيمة صغيرة لــ c مجالاً فعالاً أكبر لوسيط الاحتراس. وينفذ التصفير في F2 عندما لا يكون الانسجام بين دخل ذاكرة الأجل القصير وذاكرة الأجل الطويل المخزن قريباً بقدر كاف كما هو محدد بواسطة وسيط الاحتراس $0 < \rho < 1$ ، أي عندما:

$$\frac{\rho}{(e + \|\mathbf{r}\|)} > 1$$

يؤدي هذا الوسيط دوراً هاماً في تحديد عدد التجمعات الممكن تشكيلها في طبقة التجمع F2، ولكن قيمته المحصورة بين 0.7 و1 تكون الأهم، ولأي قيمة لهذا الوسيط أقل من 0.7 لها نفس تأثير قيمة الوسيط المساوي للصفر.

يمكن إثبات أن $\|\mathbf{r}\| = 1$ عندما ينسجم نموذج ذاكرة الأجل القصير **u** مع نموذج ذاكرة الأجل الطويل للعقدة المختارة في F2 (ولتكن العقدة J بنموذج ذاكرة الأجل الطويل $w_{Ji}$). وهكذا، وبسبب أن التصفير يحدث فقط إذا كان $\|\mathbf{r}\| \geq \rho$ فلن يحدث تصفير في F2 عندما ينسجم كثيراً **u** و$\mathbf{w}_J$ (أي، يكون $\mathrm{Cos}(\mathbf{u},\mathbf{w}_i)$ قريباً من الواحد). يحدث هذا عندما ينسجم نموذج توقع أعلى لأدنـــى مع نموذج الدخل جيداً.

يجب أن يمنع التصفير من الحدوث عندما لا توجد أية عقدة مستخدمةمنسجمة مع **u** جيداً، بعدئذ تختار عقدة غير مستخدمة من قبل لتمثيل (لتعلم) نموذج دخل جديد غير مميز. يمكن أن ينجز هذا بواسطة تخصيص أوزان العقدة غير المستخدمة بقيم أولية قريبة من الصفر (لكل العقد غير المستخدمة J، $\|\mathbf{w}_J\| \cong 0$). كما في شبكة نظرية الطنين المتكيف الأولى، يمنع التصفير مادام التعليم ضرورياً.

## 7.14 تعليم شبكة نظرية الطنين المتكيف الثانية ART2 learning

مع أن بنية شبكة نظرية الطنين المتكيف الثانية هي أكثر تعقيداً من أختها الأولى ART1، فإن عملية التعليم ستكون أساساً هي نفسها. في الحقيقة ستكون معادلات تعليم ذاكرة

الأجل الطويل أبسط نوعاً ما.

تعطى معادلات تعليم ذاكرة الأجل الطويل أدنـــى لأعلى وأعلى لأدنـــى في شبكة نظرية الطنين المتكيف الثانية على النحو التالي:

$$\frac{dv_{ij}}{dt} = g(y_j)[p_i - v_{ij}] \quad \text{أدنـــى لأعلى} \quad (27.14)$$

$$\frac{dw_{ji}}{dt} = g(y_j)[p_i - w_{ji}] \quad \text{أعلى لأدنـــى} \quad (28.14)$$

إذاً، في حالة دخل معطى، سينسجم نموذج ذاكرة الأجل القصير **u** بقدر كافٍ مع أحد نماذج ذاكرة الأجل الطويل المخزنة لعقدة F2، حيث تختار العقدة (ولتكن J) كرابح، ويحدث تعليم ما. في تلك الحالة، ومن المعادلة (24.14)، يمكـــن كتابة المعادلات (27. 14) و(28.14) كما يلي:

$$\frac{dv_{iJ}}{dt} = d(1-d)\left[\frac{u_i}{1-d} - v_{iJ}\right] \quad \text{أدنـــى لأعلى} \quad (29.14)$$

$$\frac{dw_{Ji}}{dt} = d(1-d)\left[\frac{u_i}{1-d} - w_{Ji}\right] \quad \text{أعلى لأدنى} \quad (30.14)$$

لكل عقد F2 الأخرى ($j \neq J$)، سيكون:

$$dw_{Ji}/dt = 0 \quad \text{و} \quad dv_{iJ}/dt = 0$$

وهكذا لا يحدث تعليم في هذه العقد.

لوحظ أن أوزان كل العقد غير المستخدمة في الطبقة F2 يجب أن تكون لها قيم أولية قريبة من الصفر (في حالة العقد غير المستخدمة J، $\|\mathbf{w}_J\| \cong 0$) لمنع حدوث التصفير عندما ينتج انسجام ضعيف بين **u** و$\mathbf{w}_J$.

يوضع أثر ذاكرة الأجل الطويل أدنـــى لأعلى $v_i$ بقيم أولية قريبة من الصفر، لأسباب عديدة. يمكن البرهان من المعادلة (28.14) أنه في حالة أوزان أدنـــى لأعلى الموصلة إلى العقدة رقم J في الطبقة F2، ولتكن $\mathbf{v}_J$، فإن $\|\mathbf{v}_J\| \to 1/(1-d)$ خلال التعليم. وهكذا، إذا اختيرت القيمة الأولية لـــ $\mathbf{v}_J$ أكبر من $1/(1-d)$ فسيتحول الدخل الذي يختار عقدة غير مستخدمة خلال التجربة إلى عقدة أخرى غير مستخدمة أخرى خلال هذه التجربة. ومن ثم فإنه، من الضروري وضع القيم الأولية للأوزان أدنـــى لأعلى بقيمة:

$$v_{ij}(0) \le \frac{1}{(1-d)\sqrt{n}}$$

إن اختيار قيمة كبيرة لـــ $v_{ij}(0)$ سيشجع الشبكة على تشكيل تجمعات أكثر.

ويجب أن تكون القيمة الأولية لأوزان أعلى لأدنى $w_{ji}$ صغيرة لتأكُّد عدم حدوث تصفير في النموذج الأول المتوضع على وحدة تجمع:

$$w_{ji}(0) = 0$$

سنحاول الآن تلخيص تعليم شبكة الطنين المتكيف الثانية على شكل خوارزمية بخطوات عديدة ليتسنـــى للقارئ تطبيقها مباشرة في مسائله الخاصة.

يمكن أن تستعمل هذه الخوارزمية للتعليم السريع أو للتعليم البطيء. في التعليم السريع تستمر تكرارات تغيير الأوزان وفقاً لتحديث تفعيلات F1 حتـــى الوصول إلى حالة الاستقرار. أما في التعليم البطيء، فينفَّذ تكرار واحد فقط لتغيير الأوزان وفقاً لتحديث تفعيلات F1، لكن يلزم عدد كبير من تجارب التعليم حتـــى تستقر الشبكة ( Carpenter & Grossberg عام 1987[238]). سنحاول في الأمثلة القادمة إجراء مقارنة بين نوعي التعليم. وستتكرر الحسابات التالية خلال خطوات عديدة من الخوارزمية، وسيشار لها بـــ "تحديث تفعيلات F1".

تعتبر الوحدة J هي الرابحة بعد المنافســـة، وإذا لم يتم اختيار أي وحدة رابحة فســـيكون d = 0 لجميع الوحدات. لاحظ أن حسابات $t_i$ و$p_i$ يمكن تنفيذها على التوازي، وكذلك حسابات $x_i$ و $q_i$. يعطى تحديث تفعيلات F1 بالعلاقات التالية:

$$t_i = I_i + au_i \qquad x_i = t_i/(e+\|\mathbf{t}\|)$$

$$s_i = f(x_i) + bf(q_i) \qquad u_i = s_i/(e+\|\mathbf{s}\|)$$

$$q_i = p_i/(e+\|\mathbf{p}\|) \qquad p_i = u_i + dw_{ji}$$

وتابع التفعيل المعتمد هو:

$$f(x) = \begin{cases} x & x \ge \theta \\ 0 & x < \theta \end{cases}$$

يدعى $\theta$ وسيط تخميد الضجيج، وقيمته النموذجية تساوي $1/\sqrt{n}$، وقد تكون أكبر في بعض المسائل. توضع مركبات شعاع الدخل المعياري (والأشعة الأخرى في حلقة F1) التـــي

تكون أقل من هذه القيمة مساوية للصفر.

1. إعطاء الوسطاء قيم أولية:

a, b, c, θ, d, e, α, ρ

2. كرر الخطوات من 3-13 N-EP (Number of Epochs) مرة (N-EP العدد المخصص لأدوار التدريب)

3. لكل شعاع دخل I، كرر الخطوات من 4-12

4. تحديث تفعيلات الوحدة F1:

$$u_i = 0 \qquad q_i = 0$$
$$t_i = I_i \qquad s_i = f(x_i) \qquad x_i = I_i/(e+\|\mathbf{I}\|)$$
$$p_i = 0$$

تحديث تفعيلات الوحدة F1 ثانية:

$$x_i = t_i/(e+\|\mathbf{t}\|) \qquad t_i = I_i + au_i$$
$$u_i = s_i/(e+\|\mathbf{s}\|) \qquad s_i = f(x_i) + bf(q_i)$$
$$p_i = u_i \qquad q_i = p_i/(e+\|\mathbf{p}\|)$$

5. حساب الإشارات إلى وحدات F2:

$$y_j = \sum_i v_{ij} p_i$$

6. ما دام شرط التصفير صحيحاً، كرر الخطوات من 7-8

7. إيجاد العقدة J في الطبقة F2 ذات الإشارة الأكبر (عرف J بحيث $y_J \geq y_j$، في حالة (j = 1, 2, …, m

8. افحص شرط التصفير:

$$u_i = s_i/(e+\|\mathbf{s}\|)$$
$$p_i = u_i + d\,w_{Ji}$$
$$r_i = \frac{(u_i + c\,p_i)}{e + \|\mathbf{u}\| + c\|\mathbf{p}\|}$$

إذا كان $\|\mathbf{r}\| < \rho - e$ فإن $y_J = -1$ (مخمدة)، التصفير صحيح كرر الخطوة 6

إذا كان $\|\mathbf{r}\| \geq \rho - e$ فإن:

$$t_i = I_i + au_i$$

$$x_i = t_i/(e + \|\mathbf{t}\|)$$

$$q_i = p_i/(e + \|\mathbf{p}\|)$$

$$s_i = f(x_i) + bf(q_i)$$

التصفير خطأ نفذ الخطوة 9

9. كرر الخطوات من 10-12، N-IT (Nomber of Iterations) مرة (N-IT عدد تكرارات التعليم)

10. تحديث أوزان العقدة الرابحة J:

$$w_{Ji} = adu_i + \{1 + ad(d-1)w_{Ji}\}$$

$$\upsilon_{iJ} = adu_i + \{1 + ad(d-1)w_{iJ}\}$$

11. تحديث تفعيلات F1:

$$t_i = I_i + au_i \qquad x_i = t_i/(e + \|\mathbf{t}\|)$$

$$s_i = f(x_i) + bf(q_i) \qquad u_i = s_i/(e + \|\mathbf{s}\|)$$

$$q_i = p_i/(e + \|\mathbf{p}\|) \qquad p_i = u_i + dw_{ji}$$

12. اختبار شرط توقف تحديث الأوزان

13. اختبار شرط توقف عدد الأدوار.

تذكر أن التصفير لا يحدث خلال الطنين (الخطوة 9)، ووحدة رابحة جديدة لا يمكن أن تختار خلال الطنين. عادة، في التعليم البطيء N-IT = 1 وتحذف الخطوة 11، في التعليم السريع، وفي حالة أول نموذج متعلم بواسطة التجمع، ستكون u على التوازي مع w خلال حلقة التدريب وستعطى أوزان الاستقرار بالعلاقات التالية:

$$w_{Ji} = \frac{1}{1-d} u_i$$

$$\upsilon_{iJ} = \frac{1}{1-d} u$$

**أمثلة بسيطة:**

سنعتمد الآن العديد من الأمثلة في عمل شبكة نظرية الطنين المتكيف الثانية في حالة دخل بمركبتين (n = 2). في كل هذه الأمثلة ستوضع قيم الوسطاء كما يلي:

$$A = 10, b = 10, c = 0.1, d = 0.9, e = 0$$

(لذا لن يظهر e في الصيغ القادمة، بغية التبسيط)

**مثال 6:**

في هذا المثال سيوضح أنه، للمرة الأولى ستختار وحدة التجمع الأولى كرابح، إنها ستتعلم نموذج دخل بضجيج محمّد. لن يحدث التصفير في تعلم (تعليم سريع) النموذج الأول بواسطة وحدة التجمع. وستكون الأوزان النهائية $1/(1-d)$ مرة شعاع الدخل بضجيج محمّد. وستكون قيم الوسطاء مساوية لـــ:

$\theta = 0.7$ $\rho = 0.9$

سيكون شعاع الوزن الأولي أدنـــى لأعلى لكل وحدة تجمع:

$$\mathbf{v}_j = (7.0, 7.0)$$

سيكون شعاع الوزن الأولي أعلى لأدنـــى لكل وحدة تجمع:

$$\mathbf{w}_j = (0.0, 0.0)$$

وسيكون الدخل:

$$\mathbf{I} = (0.8, 0.6)$$

وستكون جميع التفعيلات الأخرى بقيمة الصفر. في أول دورة F1، سيكون لدينا:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.8,0.6)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.8,0.0)$$

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.0,0.0)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.0,0.00)$$

$$\mathbf{p} = \mathbf{u} = (0.0,0.0)$$

في حلقة F1 الثانية:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6) + 10(1.0,0.0) = (10.8,0.6)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.998,0.055)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.998,0.0) + 10(1.0,0.0) = (10.998,0.0)$$

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (1.0,0.0)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (1.0,0.0)$$

$$\mathbf{p} = \mathbf{u} = (1.0,0.0)$$

وفي حالة تكرارات أكثر لن تتغير قيمة **u** أو **p**، لذا، سنرسل الإشارة الآن من الوحدات P إلى الطبقة F2 بحيث تجد الطبقة F2 الوحدة الفائزة. وبسبب أن هذا هو النموذج الأول المقدم إلى الشبكة، وأن أوزان أدنـــى لأعلى لكل وحدات التجمع بقيم أولية متساوية، فإن كل وحدات الطبقة F2 ستستقبل نفس الدخل. ومن ثم سنختار وحدة التجمع الأولى كرابح.

في الحلقة التـــي تختبر التصفير:

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (1.0,0.0)$$

ينفّذ اختبار التصفير فور استقبال الوحدات P إشارة أعلى لأدنـــى من وحدة التجمع الرابحة (الوحدة J). على أية حال، وبسبب أن هذه الوحدة لم تتعلم أية نماذج من قبل (ووضعت أوزان أعلى لأدنـــى بقيم أولية مساوية إلى الصفر)، فإن تفعيل الوحدات P لن يتغير بواسطة إشارة أعلى لأدنى؛ أي:

$$\mathbf{p} = \mathbf{u} + d\,\mathbf{w}_J = (1.0,0.0) + 0.9(0.0,0.0)$$

باعتبار

$$\mathbf{p} = \mathbf{u}$$

يعطي اختبار شرط التصفير:

$$\|\mathbf{r}\| = \frac{\|\mathbf{u} + c\mathbf{u}\|}{\|\mathbf{u}\| + c\|\mathbf{u}\|} = 1$$

لكي تتعلم وحدة التجمع هذه، يجب أن يكون لدينا:

$$\|\mathbf{r}\| \geq \rho$$

على كل فإن، $\|\mathbf{r}\| = 1 \geq \rho$ في حالة قيمة صحيحة لوسيط الاحتراس ρ (لأن $\rho \leq 1$)، لذا سيسمح لوحدة التجمع الرابحة بتعلم النموذج الحالي. يُظهر هذا المثال أن التصفير لن يحدث في النموذج الأول على أية وحدة تجمع.

الآن سننهي حسابات الحلقة F1:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6) + 10(1.0,0.0) = (10.8,0.6)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.998,0.055)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (1.0,0.0)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.998,0.0) + 10(1.0,0.0) = (10.998,0.0)$$

وسيتم تحديث الأوزان باستعمال معدل تعليم يساوي 0.6:

$$\mathbf{w}_J^{new} = 0.6(0.9)\mathbf{u} + [1.0 - 0.6(0.9)(0.1)]\ \mathbf{w}_J^{old}$$

$$= 0.54\ \mathbf{u} + 0.946\ \mathbf{w}_J^{old}$$

$$\mathbf{w}_J = (0.54\ ,0.0)$$

$$\mathbf{v}_J^{new} = 0.6(0.9)\mathbf{u} + [1.0 - 0.6(0.9)\ (0.1)]\ \mathbf{v}_J^{old}$$

$$= 0.54\ \mathbf{u} + 0.946\ \mathbf{v}_J^{old}$$

$$\mathbf{v}_J = (0.54,\ 0.0) + (6.62,\ 6.62)$$

$$= (7.16,\ 6.62)$$

في الحلقة الأولى:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6) + 10(1.0,0.0) = (10.8,0.6)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.998,0.055)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.998,0.0) + 10(1.0,0.0) = (10.998,0.0)$$

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (1.0,0.0)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (1.0,0.00)$$

$$\mathbf{p} = \mathbf{u} + d\mathbf{w}_J = (1.0,0.0) + 0.9(0.09,0.0)$$

تحديث الأوزان ثانية:

$$\mathbf{w}_J^{new} = 0.6(0.9)\ \mathbf{u} + [1.0 - 0.6(0.9)(0.1)]\ \mathbf{w}_J^{old}$$

$$= 0.54\ \mathbf{u} + 0.946\ \mathbf{w}_J^{old}$$

$$\mathbf{w}_J = (0.54\ ,\ 0.0) + 0.946\ (0.54\ ,\ 0.0)$$

$$= (1.05, 0.0)$$

$$\mathbf{v}_J^{new} = 0.6(0.9)\mathbf{u} + [1.0 - 0.6(0.9)(0.1)]\,\mathbf{v}_J^{old}$$

$$= 0.54\,\mathbf{u} + 0.946\,\mathbf{v}_J^{old}$$

$$\mathbf{v}_J = (0.54, 0.0) + (6.77, 6.26)$$

$$= (7.32, 6.26)$$

وفي الحلقة F1:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6) + 10(1.0,0.0) = (10.8,0.6)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.998,0.055)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.998,0.0) + 10(1.0,0.0) = (10.998,0.0)$$

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (1.0,0.0)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (1.0,0.00)$$

$$\mathbf{p} = \mathbf{u} + d\mathbf{w}_J = (1.0,0.0) + 0.9(0.18,0.0)$$

نلاحظ أن $\mathbf{p}$ لم يكسب مطلقاً توزيعاً للمركبة الصفرية، و$\mathbf{q}$ و$\mathbf{u}$ لم يتغيرا، وينمو $\mathbf{w}_J$ تدريجياً إلى ضعف $\mathbf{u}$. في الحقيقة، وبسبب أن $\mathbf{u}$ لم يتغير خلال التعليم، يمكن أن توجد القيم المستقرة للأوزان مباشرة من الصيغ التالية:

$$\frac{d}{dt} w_{Ji} = du_i + d(d-1)w_{Ji}$$

$$0 = du_i + d(d-1)w_{Ji}$$

$$w_{Ji} = \frac{1}{1-d}u_i = \frac{1}{0.1}\cdot(1.0) = 10(1.0) = (10.0)$$

$$\mathbf{w}_J = (10\ ,\ 0)$$

ومع أن أوزان أدنى لأعلى بدأت من قيم أولية مختلفة عن أوزان أعلى لأدنى، ولنفس المعادلات التفاضلية، فإن هذه الأوزان تقاربت لنفس القيم. وهكذا، نجد أن أوزان الاستقرار للنموذج الأول المتعلم بواسطة أية وحدة تجمع يمكن أن يوجد بدون حل تكراري للمعادلات التفاضلية للأوزان.

هناك ميزتان خاصتان لهذا المثال، هما أن النموذج هو الأول المتعلم بواسطة وحدة التجمع

وبعض مركبات الدخل كانت مخمدة. الصيغة الأساسية لشبكة نظرية الطنين المتكيف الثانية تقترح أن $\theta = 1/\sqrt{n}$، التـــي تعطي قيمة تقريبية 0.7 في حالة n = 2. على أية حال، من السهل رؤية أن استعمال هذه القيمة لـــ θ سيقود أي دخل (مركباته ليست متساوية تماماً) إلى $\mathbf{u} = (1,0)$ أو $\mathbf{u} = (0,1)$ بعد التكرار الأول. وهكذا، نرى أن اختيار وسيط الضجيج θ يستطيع التأثير بقوة في إنجاز الشبكة.

**مثال 7:**

سنوضح في هذا المثال تأثير أوزان أدنـــى لأعلى الأولية في عدد التجمعات المشكلة باستعمال التعليم السريع. إن وسيط تخميد الضجيج وشعاع الدخل هما:

$$\theta = 0.1 \qquad \mathbf{I} = (0.8\,,\,0.6)$$

وستعطى جميع التفعيلات الأخرى قيماً أولية مساوية للصفر.

ففي الحلقة F1 الأولى، لدينا:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6)$$
$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.8,0.6)$$
$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.8,0.6)$$
$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.0,0.0)$$
$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.0,0.00)$$
$$\mathbf{p} = \mathbf{u} = (0.0\,,0.0)$$

وفي الحلقة F1 الثانية:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6) + 10(0.8,0.6) = (8.8,6.6)$$
$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.8,0.6)$$
$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.8,0.6) + 10(0.8,0.6) = (8.8,6.6)$$
$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.8,0.6)$$
$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.8,0.6)$$
$$\mathbf{p} = \mathbf{u} = (0.8,0.6)$$

وفي حالة تكرارات أكثر لن يحدث تغير في **u** أو **p**، لذا سيبدأ تكرار الدورة F2-F1

لإيجاد الرابح:

- ترسل الإشارات من الوحدات P إلى الطبقة F2.
- تتنافس وحدات الطبقة F2 غير المخمدة من قبل لإيجاد الوحدة الرابحة.
- تـــرسل وحدة F2 الرابحة إشارة خرجها عكسياً إلى الطبقة F1، لأن هذا هـــو النموذج الأول المتعلم بواسطة هذا التجمع (وتساوي أوزان أعلى لأدنـــى الأولية الصفر).

بوجه عام، سنحدث تفعيلات الوحدات P لدمج إشارة أعلى لأدنـــى من وحدة F2 الرابحة، ومن ثم فحص شرط التصفير. إذا تحقق الشرط، سنحدث تصفير تفعيلات F1. على أية حال، كما في المثال 6 السابق، لن يحدث التصفير في النموذج الأول على التجمع إذا كانت أوزان أعلى أدنـــى الأولية بقيمة تساوي الصفر. لدينا:

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.8,0.6)$$

$$\mathbf{p} = \mathbf{u} + d\mathbf{w}_J = (0.8,0.6)$$

سيجري اختبار شرط التصفير عند هذه النقطة.

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.8,0.6) + 10(0.8,0.6) = (8.8,6.6)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.8,0.6)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.8,0.6) + 10(0.8,0.6) = (8.8,6.6)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.8,0.6)$$

تحديث الأوزان، باستعمال معدل التعليم بقيمة 0.6:

$$w_{Ji}^{new} = 0.6(0.9)u_i + [1.0 - 0.6(0.9)(0.1)]\ w_{Ji}^{old}$$

$$= 0.54\ u_i + 0.946\ w_{Ji}^{old}$$

$$\mathbf{w} = 0.54(0.8\ ,\ 0.6)$$

$$= (0.432,\ 0.324)$$

تحديث F1:

$$\mathbf{u} = (0.8,\ 0.6)$$

$$\mathbf{t} = (0.8,\ 0.6) + 10(0.8,\ 0.6)$$

$$= 11(0.8,\ 0.6)$$

$\mathbf{p} = (0.8, 0.6) + 0.9(0.432, 0.324)$

$= 1.486(0.8, 0.6)$

$\mathbf{x} = (0.8, 0.6)$

$\mathbf{q} = (0.8, 0.6)$

$\mathbf{s} = (0.8, 0.6) + 10(0.8, 0.6)$

$=11(0.8, 0.6)$

الآن سنحدث أوزان أعلى لأدنى:

$\mathbf{w} = 0.54\ \mathbf{u} + 0.946\ \mathbf{w}^{old}$

$= 0.54(0.8, 0.6) + 0.946(0.54)\ (0.8, 0.6)$

$= 1.05(0.8, 0.6)$

باعتبار كل الأشعة من مضاعفات (0.8, 0.6)، من السهل أن نرى أن أوزان أعلى لأدنـــى ستتقارب إلى مضاعقات (0.8, 0.6). في الحقيقة، باعتبار أن **u** ثابت، تعرف أوزان أعلى لأدنـــى وأدنـــى لأعلى المستقرة من الصيغ التالية:

$$\frac{d}{dt} w_{Ji} = du_i + d(d-1)w_{Ji}$$

$$0 = du_i + d(d-1)w_{Ji}$$

$$w_{Ii} = \frac{1}{1-d} u_i$$

$\mathbf{w}_1 = 10(0.8, 0.6) = (8.0, 6.0)$

$\mathbf{v}_1 = 10(0.8, 0.6) = (8.0, 6.0)$

سنتابع المثال بتقديم النموذج الثاني. الآن ستكون أوزان أدنـــى لأعلى كما يلي:

$\mathbf{v}_1 = (8.0, 6.0)$ $\mathbf{v}_2 = (7.0, 7.0)$

وستكون أوزان أعلى لأدنـــى كما يلي:

$\mathbf{w}_1 = (8.0, 6.0)$ $\mathbf{w}_2 = (0.0, 0.0)$

وسنقدم نموذج الدخل الثانـــي التالي:

$\mathbf{I} = (0.6, 0.8)$

وجميع التفعيلات الأخرى ستكون مساوية للصفر.

في الدورة F1 الأولى، سيكون لدينا:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.6,0.8)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.6,0.8)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.6,0.8)$$

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.0,0.0)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.0,0.00)$$

$$\mathbf{p} = \mathbf{u} = (0.0,0.0)$$

وفي الدورة F1 الثانية، سيكون لدينا:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.6,0.8) + 10(0.6,0.8) = (6.6,8.8)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.6,0.8)$$

$$\mathbf{s} = f(\mathbf{x}) + bf(\mathbf{q}) = (0.6,0.8) + 10(0.6,0.8) = (6.6,8.8)$$

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.6,0.8)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.6,0.8)$$

$$\mathbf{p} = \mathbf{u} = (0.6,0.8)$$

لن تغير تكرارات أكثر **u** أو **p**، لذا، فإن تكرارات الدورة F2-F1 لإيجاد الوحدة الرابحة يمكن أن تبدأ.

ترسل الإشارات من الوحدات P إلى الطبقة F2.

سيكون دخل الشبكة net إلى وحدة التجمع الأولى:

$$(0.6,\ 0.8)\ (8.0,\ 6.0) = 4.8 + 4.8 = 9.6$$

سيكون دخل الشبكة net إلى وحدة التجمع الثانية:

$$(0.6,\ 0.8)\ (7.0\ ,\ 7.0) = 4.2 + 5.6 = 9.8$$

وهكذا، ستكون الوحدة الرابحة هي الوحدة الثانية. وستتعلم وحدة التجمع الثانية هذا النموذج بأسلوب مشابه لما سبق وصفه في تعلم النموذج الأول بواسطة وحدة التجمع

الأولى.

على أية حال، إذا أعطيت أوزان أدنـــى لأعلى الأولية القيمة (5.0, 5.0)، عوضاً عن القيمة المسموحة الأعظمية (7.0, 7.0)، فإن النموذج الثانـــي سيختار التجمع الأول كرابح. وستحدد قيمة وسيط الاحتراس إذا كان التجمع الأول سيتعلم هذا النموذج.

**مثال 8:**

سنتابع المثال 7 السابق باستعمال قيمة منخفضة بقدر كافٍ لوسيط الاحتراس بحيث تتعلم وحدة التجمع الأول النموذج الثاني. وهكذا، يقدم النموذج الأول (0.8, 0.7) إلى الشبكة، وستكون أوزان أعلى لأدنـــى وأدنـــى لأعلى في الوحدة الأولى للتجمع (8.0, 6.0). يقدم النموذج الثانـــي (0.6, 0.8) إلى الشبكة، وتنفذ تكرارات الحلقة F1 كما في المثال 7 السابق. تستمر الحسابات لتعيين وحدة F2 الرابحة.

سيكون دخل الشبكة لوحدة التجمع الأولى:

$$(0.6, 0.8)\,(8.0, 6.0) = 4.8 + 4.8 = 9.6$$

سيكون دخل الشبكة لوحدة التجمع الثانية:

$$(0.6, 0.8)\,(5.0, 5.0) = 3.0 + 4.0 = 7.0$$

الآن، ستكون وحدة التجمع الأولى هي الرابحة. سترسل الوحدة الرابحة إشارة أعلى لأدنـــى وسيفحص شرط التصفير:

$$\mathbf{u} = \mathbf{s}/\|\mathbf{s}\| = (0.6, 0.8)$$

$$\mathbf{p} = \mathbf{u} + d\mathbf{w}_J = (0.6, 0.8) + 0.9(8.0, 6.0) = (7.8, 6.2)$$

يتطلب شرط التصفير تحقق الشرط التالي كي يسمح لهذه الوحدة بالتعليم:

$$\|\mathbf{r}\| = \frac{\|\mathbf{u} + c\mathbf{p}\|}{\|\mathbf{u}\| + c\|\mathbf{p}\|} > \rho$$

في هذه الحالة:

$$\begin{aligned} \mathbf{u} + c\mathbf{p} &= (0.6, 0.8) + (0.78, 0.62) \\ &= (1.38, 1.42) \\ \|\mathbf{u} + c\mathbf{p}\| &= 1.98 \end{aligned}$$

$$\|\mathbf{p}\| = 9.964$$

$$\|\mathbf{u}\| + 0.1\|\mathbf{p}\| = 1.9964$$

$$\|\mathbf{u}\| = 0.992 > 0.9 = \rho$$

الوحدة الرابحة مقبولة، لذا، سيحدث فإن تصفير تفعيلات F1:

$$\mathbf{t} = \mathbf{I} + a\mathbf{u} = (0.6\ ,0.8) + 10(0.6\ ,0.8) = (6.6\ ,8.8)$$

$$\mathbf{x} = \mathbf{t}/\|\mathbf{t}\| = (0.6\ ,0.8)$$

$$\mathbf{q} = \mathbf{p}/\|\mathbf{p}\| = (0.78\ ,0.62)$$

ومن ثم يجري تحديث الأوزان باستعمال معدل تعليم يساوي 0.6:

$$w_{Ji}^{new} = 0.6(0.9)\mathrm{u}_i + [1.0 - 0.6(0.9)\ (0.1)]\ w_{Ji}^{old}$$

$$= 0.54\ \mathrm{u}_i + 0.946\ w_{Ji}^{old}$$

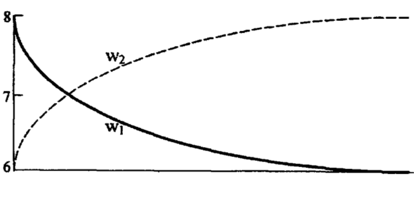
$$\mathbf{w} = 0.54(0.6\ ,\ 0.8) + 0.946(8.0\ ,\ 6.0)$$

$$= (7.9\ ,\ 6.1)$$

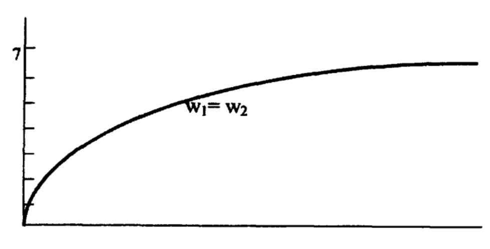
خلال التعليم تتبع الدورة F1 كل تحديث وزن. يوضح (الشكل 5.14) تطور الأوزان. لاحظ أن الأوزان النهائية هي في الأساس شعاع الدخل الثاني؛ عملياً ضاعت كل المعلومات حول الشعاع الأول المتعلم بواسطة هذا التجمع .

**مثال 9:**

إذا كررنا المثال 8 السابق باستعمال التعليم البطيء، سنرى أن شعاع الوزن في نهاية الأمر يصبح متوسط النموذجين المتوضعين على التجمع. هذا موضح في (الشكل 6.14).




**الشكل 5.14:** تغير الأوزان للمثال 8




**الشكل 6.14**: تغير الأوزان للمثال 9

**مثال 10**:

يوضح هذا المثال حقيقة أنه باستعمال التعليم البطيء، وبعد تدريب تجمع متعلم من قبل، ستكون الأوزان متوسط النماذج المتوضعة على ذلك التجمع. يوضح (الشكل 7.14) تأثير تقديم الشعاع الأول:

(0.2, 0.4, 0.6, 0.8, 1.0)

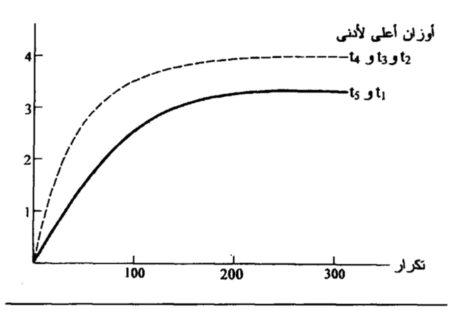
متبوعاً بالشعاع الثاني:

(1.0, 0.8, 0.6, 0.4, 0.2)

اختيرت وسطاء الشبكة بحيث يتوضع كل من الشعاعين على نفس وحدة التجمع (في هذه الحالة، لدينا فقط وحدة تجمع واحدة؛ سينجز نفس التأثير باختيار أوزان أدنى لأعلى أولية صغيرة بقدر كافٍ). وضع وسيط تخميد الضجيج بحيث تكون المركبة الأصغر في كل شعاع دخل فقط مخمدة.

**مثال 11**:

يمكن استعمال عينة من المعطيات المقترحة من قبل Kohanen عام 1989[227] لتوضيح سلوك الشبكات العصبونية لنظرية الطنين المتكيف الثانية.




**الشكل 7.14**: تغير الأوزان للمثال 10

يمكن توضيح العلاقات بين النماذج تخطيطياً كما في (الشكل 8.14). كما ذكرنا من قبل، تختلف النماذج المعروضة في سطر أو عمود بعضها عن بعض بمركبة واحدة فقط. وكذلك، توافق المسافة بين النماذج في نفس السطر أو العمود على مخطط (الشكل 8.14) مباشرة المسافة الإقليدية بين الشعاعين. مثلاً، تختلف النماذج X وY المتجاورة فقط في المركبة الرابعة والمسافة الإقليدية بين (0 ,2 ,6 ,3 ,3) و(0 ,3 ,6 ,3 ,3) تساوي الواحد. المعطيات الأصلية معطاة في (الشكل 9.14).

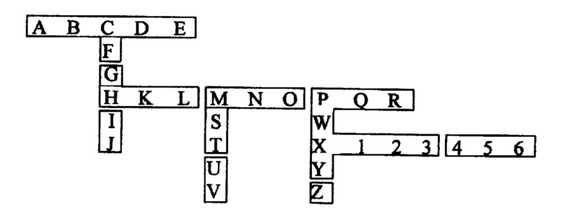
| *A* | *B* | *C* | *D* | *E* | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *F* | | | | | | | | | | | | |
| | | *G* | | | | | | | | | | | | |
| | | *H* | *K* | *L* | *M* | *N* | *O* | *P* | *Q* | *R* | | | | |
| | | *I* | | | *S* | | | *W* | | | | | | |
| | | *J* | | | *T* | | | *X* | 1 | 2 | 3 | 4 | 5 | 6 |
| | | | | | *U* | | | *Y* | | | | | | |
| | | | | | *V* | | | *Z* | | | | | | |

**الشكل 8.14**: بنية معطيات اختبار شجرة العبور

| | | | | | |
|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | A |
| 2 | 0 | 0 | 0 | 0 | B |
| 3 | 0 | 0 | 0 | 0 | C |
| 4 | 0 | 0 | 0 | 0 | D |
| 5 | 0 | 0 | 0 | 0 | E |
| 3 | 1 | 0 | 0 | 0 | F |
| 3 | 2 | 0 | 0 | 0 | G |
| 3 | 3 | 0 | 0 | 0 | H |
| 3 | 4 | 0 | 0 | 0 | I |
| 4 | 5 | 0 | 0 | 0 | J |
| 3 | 3 | 1 | 0 | 0 | K |
| 3 | 3 | 2 | 0 | 0 | L |
| 3 | 3 | 3 | 0 | 0 | M |
| 3 | 3 | 4 | 0 | 0 | N |
| 3 | 3 | 5 | 0 | 0 | O |
| 3 | 3 | 6 | 0 | 0 | P |
| 3 | 3 | 7 | 0 | 0 | Q |
| 3 | 3 | 8 | 0 | 0 | R |
| 3 | 3 | 3 | 1 | 0 | S |
| 3 | 3 | 3 | 2 | 0 | T |
| 3 | 3 | 3 | .3 | 0 | U |
| 3 | 3 | 3 | 4 | 0 | V |
| 3 | 3 | 6 | 1 | 0 | W |
| 3 | 3 | 6 | 2 | 0 | X |
| 3 | 3 | 6 | 3 | 0 | Y |
| 3 | 3 | 6 | 4 | 0 | Z |
| 3 | 3 | 6 | 2 | 1 | 1 |
| 3 | 3 | 6 | 2 | 2 | 2 |
| 3 | 3 | 6 | 2 | 3 | 3 |
| 3 | 3 | 6 | 2 | 4 | 4 |

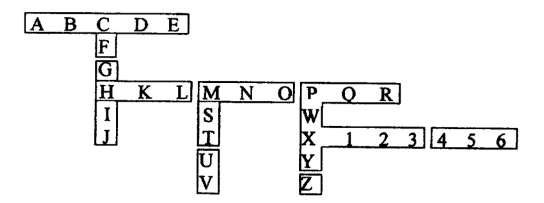
**الشكل 9.14**: معطيات اختبار شجرة العبور

يوضح (الشكل 11.14) التجمعات المشكلة باستعمال معطيات شبكة العبور والتعليم السريع، ويوضح (الشكل 10.14) نتائج التعليم البطيء. ستكون قيم وسيط الاحتراس وتخميد الضجيج عالية نسبياً، $\rho = 0.99$ و$\theta = 0.447$.




**الشكل 10.14:** تجمعات معطيات اختبار شجرة العبور باستعمال التعليم السريع

($\theta = 0.447, \rho = 0.99$)




**الشكل 11.14:** تجمعات معطيات اختبار شجرة العبور باستعمال التعليم البطيء

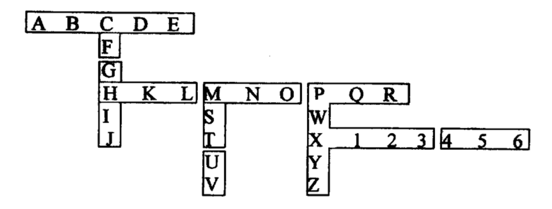
($\theta = 0.447, \rho = 0.99$)

**مثال 12:**

يوضح (الشكلان 12.14 و13.14) التجمعات المتشكلة باستعمال التعليم السريع والبطيء في حالة وسيط احتراس بقيمة معتدلة مع وسيط تخميد ضجيج عال، $\rho = 0.95$ و$\theta = 0.447$. لاحظ أن الشبكة كانت حساسة للتغيرات الصغيرة في قيمة وسيط الاحتراس. وكما هو متوقع ، ومرغوب أيضاً، فقد شكّلت تجمعات أكثر في حالة وسيط احتراس عالٍ.

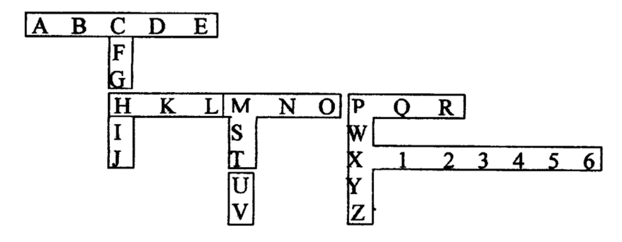
| | التجمعات المتشكلة | |
|---|---|---|
| | $\rho = 0.95$ | $\rho = 0.99$ |
| التعليم السريع | 6 | 8 |
| التعليم البطيء | 4 | 7 |

ففي كل من هذه الحالات جرى تنفيذ 1000 دور، مع تكرار تحديث وزن واحد لكل دور في التعليم البطيء، ونفذت ثلاثة أدوار تدريب في التعليم السريع.




**الشكل 12.14:** تجمعات معطيات اختبار شجرة العبور باستعمال التعليم السريع

($\theta = 0.447$ و $\rho = 0.95$)



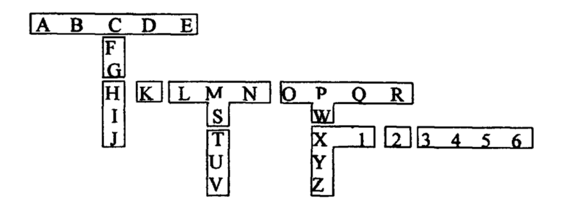

**الشكل 13.14:** تجمعات معطيات اختبار شجرة العبور باستعمال التعليم البطيء

($\theta = 0.447$ و $\rho = 0.95$)

**مثال 13:**

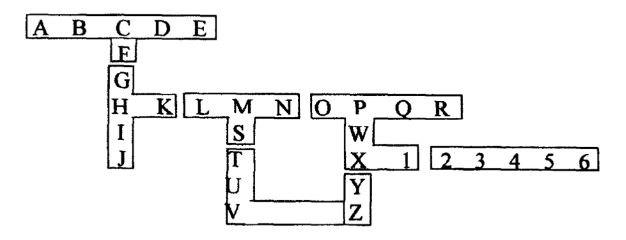
عندما يعطى وسيط الاحتراس قيمة عالية وتعطى قيمة معتدلة لوسيط كبت الضجيج $\rho = 0.99$ و$\theta = 0.2$، سيكون هناك اختلاف كبير في عدد التجمعات المشكلة باستعمال التعليم السريع والبطيء كما هو موضح في (الشكلين 14.14 و15.14)، إذا قورنت مع النتائج في حالة وسيط تخميد ضجيج عال (المثال11).

في كل من هذه الأمثلة سمح للشبكة بعدد أعظمي من عقد التجمع حدد بعشر عقد. استَعمل التعليم السريع كل العقد العشر، أما التعليم البطيء فقد استَعمل ست عقد فقط.




**الشكل 14.14**: تجمعات معطيات اختبار شجرة العبور باستعمال التعليم البطيء

($\theta = 0.2$ و $\rho = 0.99$)




**الشكل 15.14**: تجمعات معطيات اختبار شجرة العبور باستعمال التعليم البطيء

($\theta = 0.2$ و $\rho = 0.99$)

**مثال 14:**

يمكن أن تستعمل شبكة نظرية الطنين المتكيف الثانية بمعطيات ثنائية. باستعمال التعليم البطيء يمكن الحصول على أوزان مفيدة أكثر كنماذج من تلك المشكلة باستعمال شبكة نظرية الطنين المتكيف الأولى (المثال 3). ستكون التجمعات المشكلة أقل حساسية لترتيب التمثيلات مقارنة مع تلك التـــي في المثال 3. في هذا المثال سنعتمد كأشعة دخل تمثيلات

الأحرف السبعة من تشكيلات مختلفة المعطاة في المثال 3 في (الشكل 3.14).

أعطيت النتائج الموضحة في (الأشكال 16.14 و17.14) باستعمال القيم القياسية للوسطاء التالية: $a = 10$، و$b = 10$، و$c = 0.1$، و$d = 0.9$ مع وسيط احتراس يساوي 0.8، وسيط تخميد ضجيج يساوي 0.126.

إذا كان ترتيب التمثيلات على النحو التالي:

$A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3, D_1, D_2, D_3, E_1, E_2, E_3, J_1, J_2, J_3, K_1, K_2, K_3$

فإن النماذج ستكون موضوعة في التجمعات كما يلي:

| النماذج | التجمع |
| --- | --- |
| $A_1, A_2$ | 1 |
| $A_3$ | 2 |
| $C_1, C_2, C_3, D_2$ | 3 |
| $B_1, D_1, E_1, K_1, B_3, D_3, E_3, K_3$ | 4 |
| $K_2$ | 5 |
| $J_1, J_2, J_3$ | 6 |
| $B_2, E_2$ | 7 |

وستكون الأوزان من أجل التجمعات المشكلة بالترتيب الأول للتمثيلات على الشكل التالي:

التجمع1 التجمع2 التجمع3 التجمع4 التجمع5 التجمع6 التجمع7

**الشكل 16.14**: أوزان أعلى لأدنى ثنائية البعد

وإذا كانت المعطيات مرتبة على النحو التالي:

$A_1, B_1, C_1, D_1, E_1, J_1, K_1, A_2, B_2, C_2, D_2, E_2, J_2, K_2, A_3, B_3, C_3, D_3, E_3, J_3, K_3$

فإن النتائج ستكون كما يلي:

| التجمع | النماذج |
|---|---|
| 1 | $A_1, A_2$ |
| 2 | $B_1, D_1, E_1, K_1, B_3, D_3, E_3, K_3$ |
| 3 | $J_1, J_2, J_3$ |
| 4 | $B_1, D_1, E_1, K_1, B_3, D_3, E_3, K_3$ |
| 5 | $B_2, D_2, E_2$ |
| 6 | $K_2$ |
| 7 | $A_3$ |

وستكون الأوزان للتجمعات المشكلة بالترتيب الثانـــي للتمثيلات على الشكل التالي:

| التجمع1 | التجمع2 | التجمع3 | التجمع4 | التجمع5 | التجمع6 | التجمع7 |
|---|---|---|---|---|---|---|
| 00##000 | ####### | 00###0# | 0000### | ######0 | #0000#0 | 000#000 |
| 000#000 | 0#000## | 0#000## | 00000#0 | #000000 | #000#00 | 000#000 |
| 000#000 | 0#00000 | #00000# | 00000#0 | #00000# | #00#000 | 00#0#00 |
| 00#0#00 | 0#0#000 | #000000 | 00000#0 | #00000# | #0#0000 | 00#0#00 |
| 00#0#00 | 0##0000 | #000000 | 00000#0 | #####00 | ##00000 | 0#000#0 |
| 0#####0 | 0#00000 | #000000 | 00000#0 | #00000# | #0#0000 | 0#####0 |
| 0#####0 | 0#00000 | #00000# | 0#000#0 | #00000# | #00#000 | #00000# |
| ##000## | 0#000## | 0#000#0 | 0#000#0 | #000000 | #000#00 | #00000# |
| ###0### | ####### | 00###00 | 00###00 | ######0 | #0000#0 | ##000## |

**الشكل 17.14** أوزان أعلى لأدنى ثنائية البعد

لاحظ أنه على الرغم من اختلاف ترتيب التجمعات في ترتيبَـــي التمثيلات، فإن النماذج متوضعة توضعاً متطابقاً كثيراً. يمكن تمثيل أوزان أعلى لأدنى في شكل مصفوفة ثنائي البعد (تماماً مثل معطيات الدخل الأصلية الممثلة بنموذج ثنائي البعد). ستكون الأوزان إما بقيمة الصفر (المشار إليها برمز "0") وإما بقيمة متوسط إشارات الدخل للنماذج المتوضعة على التجمع (المشار إليها برمز "#").

## 8.14 شبكات نظرية الطنين المتكيف الأخرى

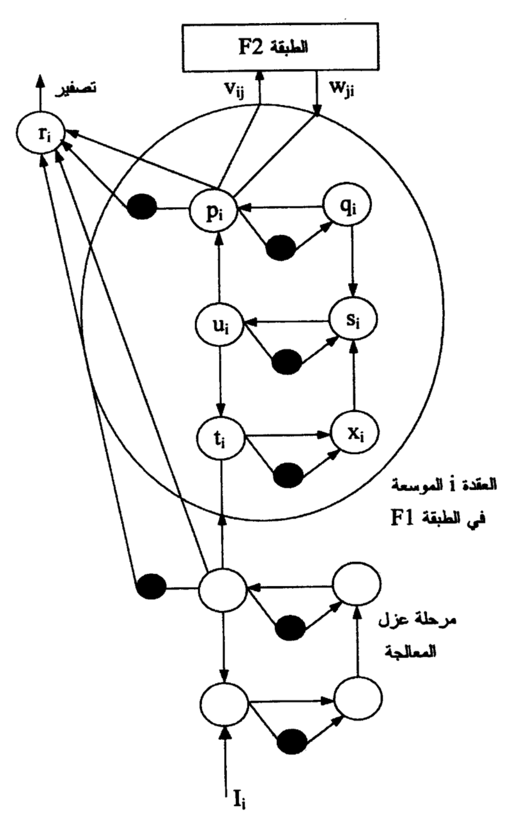
## Other ART networks

أجريت بعض التعديلات على شبكات نظرية الطنين المتكيف الأساسية في تطبيقات خاصة، أو للحصول على استقرار إضافي. مثلاً، بإضافة مرحلة عزل المعالجة إلى الشبكة، فإن شعاع الدخل **I** يمكن أن يستعمل مباشرة كدخل لنظام التوجيه الجزئي بدلاً من الشعاع **u**. تقلد طبقة العزل (تشابه) حلقات الأعلى والأدنى للطبقة F1، كما هو موضح في (الشكل 18.14).

من محاسن هذا التعديل أن **I** لا يتغير عندما تصبح F2 فعالة، وبذلك نعطي دخلاً مستقراً أكثر لنظام التوجيه الجزئي خلال دور العمل. وضعت التفاصيل الكاملة لهذا التعديل ولتعديلات أخرى من قبل Grossberg وCarpenter عام 1987[237].

هناك تعديل آخر أكثر أهمية على شبكة نظرية الطنين المتكيف هو ARTMAP العائم (Carpenter عام 1991[139]). وسمّي بالعائم لاستخدام المنطق العائم في عملية التدريب. سنعطي نسخة مبسطة عن هذا التعديل لشبكة نظرية الطنين المتكيف بدخل ذي قيم حقيقية، أسس على التعليم بمعلم ( Kasuba عام 1993[193]).

البنية الأساسية هي شبكة نظرية الطنين المتكيف الثانية المبسطة بطبقتين إضافيتين: طبقة مرمز المتمم CC (Complement Coder)، وطبقة الفئة (أو التجمع)، كما هو موضح في (الشكل 19.14). يأخذ مرمز المتمم شعاع ملمح الدخل ذي البعد n، حيث جعلت كل مركبة معيارية بقيمة محصورة ضمن المجال [0, 1]، ويعطي خرجاً ببعد 2n؛ يتألف شعاع الخرج من شعاع الدخل ومتمم هذا الشعاع على الترتيب ( n مركبة لشعاع الدخل الأصلي + n مركبة لمتممه). مثلاً، إذا كان الدخل الأصلي غير المُتَمَّم هو شعاعاً ثلاثي البعد [0.2, 0.5, 0.7]، فإن خرج مرمز المتمم إلى طبقة فئة الخرج OC (Output Category) سيكون شعاعاً سداسي البعد مؤلفاً من المداخل الثلاثة الأصلية إضافة إلى متمماتها المرتبة على التتالي:




**الشكل 18.14**: شبكة نظرية الطنين المتكيف الثانية بدخل معزول في الطبقة F1

[0.3 , 0.5 , 0.8 , 0.7 , 0.5 , 0.2]. أضيفت مركبات الدخل الأصلية (غير المتممة) لجعل الشبكة تشكل مناطق الاستقرار بسهولة.

لاحظ أن مجموع مركبات خرج مرمز المتمم هو تماماً بعد الدخل n، أي $\sum_i I_i + \sum_i (1 - I_i) = n$. استعملت هذه الخاصية للمرمز المتمم لمعايرة أشعة سطر معطيات الدخل. تحتوي طبقة الفئة العليا m عقدة، وهو العدد الأعظمي لفئات الشبكة التـــي تستطيع تعلمها. كل منها معنون (مؤشر بدليل) كفئة أو صف مفرد. تعمل هذه الطبقة خزاناً لتخزين الفئات المختلفة (m فئة) التـــي تستطيع الشبكة تعلمها.

فهي تستقبل معلومات دخل فئة فقط خلال عملية التعليم بمعلم وتعطي خرج فئة معنونة مفردة لكل نموذج دخل معطى، كما حدد بواسطة تصنيف طبقة فئة الخرج OC لنموذج الدخل.

عندما يقدم الدخل **I** إلى الشبكة، تصبح عقدة طبقة فئة الخرج J التـــي سيكون لها قيمة التفعيل الأكبر $T_J$ هي الرابحة، $T_J = \max_j \{T_j\}$، في مسابقة التعليم التنافسي. بنيت مقارنة التفعيل الأعظمي ، فعلياً، على الروابط العائمة لأشعة الدخل وأشعة أوزان عقدة فئة الخرج (راجع الفقرة (6.3) من الفصل الثالث). يعطى تفعيل العقدة رقم j بما يلي:

$$T_j(\mathbf{I}, \mathbf{W}) = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{|\mathbf{W}_j| + e} \qquad (31.14)$$

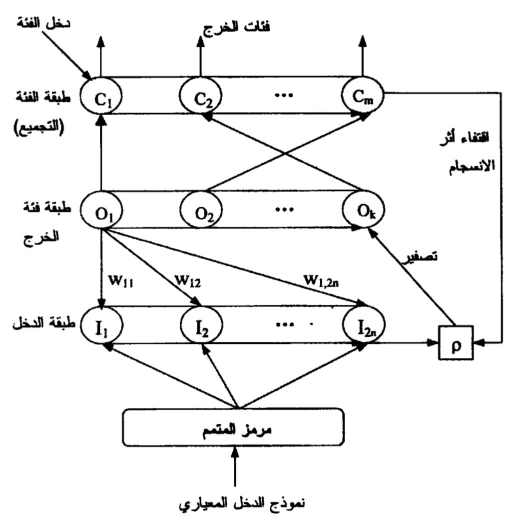
حيث نستعمل العلاقة:

$$|\mathbf{I} \wedge \mathbf{W}_j| = \sum_i \left(\min_j \{I_i, w_{ji}\}\right)$$

ويؤخذ المجموع عبر القيم الدنيا لأزواج الدخل ومركبات أشعة الوزن الموافقة. الثابت e عدد موجب صغير ، $0 < e << 1$، واستعمل حد المقام $\|\mathbf{W}_j\| = \sum_i w_{ji}$ كعامل معايرة.

يشبه تابع الانسجام $\mathbf{M}_j$ تابع تفعيل عقدة فئة الخرج، لكن جرت معايرته بالمقام $\|\mathbf{I}\| = n$. فهو يقيس درجة انتماء خرج مرمز المتمم **I** كمجموعة جزئية عائمة في $\mathbf{W}_j$، وتعطى أوزان العقدة رقم j لفئة الخرج كما يلي:

$$\mathbf{M}_j(\mathbf{I}, \mathbf{W}) = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{|\mathbf{I}| + e} = \frac{|\mathbf{I} \wedge \mathbf{W}_j|}{n} \tag{23.14}$$




الشكل 19.14: **ARTMAP** العائم البسيط

تقارن قيمة تابع الانسجام للعقدة الرابحة في فئة الخرج مع وسيط الاحتراس $\rho$ ويقدح التصفير (يتفعل) متـــى كان الانسجام بين شعاع الدخل و(أوزان) ذاكرة الأجل الطويل لعقدة فئة الخرج المختارة ليس قريباً القرب الكافي المنشود، أي:

$$\frac{|\mathbf{I} \wedge \mathbf{W}_j|}{n} < \rho \tag{33.14}$$

كإجراءات التعليم بمعلم الأخرى، هناك طوران للعملية: طور التعليم وطور التصنيف،

وسنصف أولاً طور التعليم. تتألف مجموعة التعليم من أزواج فئة/نموذج دخل. خلال التعليم، عندما يعطى نموذج فئة جديدة لدخل الشبكة، فإن عقدة خرج جديدة تكون مستخدمة في طبقة فئة الخرج وذلك بوضع أوزانها مساوية لنموذج الدخل **I** (مساوية لخرج مرمز المتمم) وتوصيل خرج العقدة إلى عقدة فئة معنونة عنونة مناسبة.

إذا كان نموذج الدخل معروفاً (معنوناً بفئة عقدة خرج مستخدمة)، تجري مقارنة الانسجام بين أوزان العقدة الرابحة في فئة الخرج (لعقدة ذات القيمة $T_j$ الأكبر) والدخل (معادلة (32.14)). تنفذ مقارنة الانسجام بقيمة وسيط اليقظة الأولي $\rho_0$. إذا كان هناك عدم انسجام (تتحقق المعادلة (33.14))، ويحدث تصفير طبقة فئة الخرج لمنع الرابح خلال مدة التدريب ولا يحدث تعليم. أيضاً، تصبح العقدة ذات $T_j$ الأكبر من بين العقد غير المخمدة المتبقية (التـــي لم يجرِ لها تصفير من قبل) فائزة جديدة.

تستمر العملية حتـــى لا يحدث أي تصفير أو تكون جميع العقد المستخدمة مخمدة. إذا لم يوجد أي عقدة بانسجام قريب بقدر كاف كما هو محدد بوسيط الاحتراس، تختار عقدة غير مستخدمة جديدة في طبقة فئة الخرج لتمثل (ترميز) الدخل **I** كفئة جديدة.

خلال التدريب، إذا شوهدت فئة الدخل ووجدت عقدة بقيمة انسجام أكبر أو تساوي قيمة وسيط الاحتراس (لا يحدث تصفير)، ولكن الفئة لا تنسجم مع نموذج التدريب المعنون، في هذه الحالة، تستخدم عقدة فئة جديدة كما وصف آنفاً. إضافة إلى ذلك، يكون وسيط اليقظة $\rho$ مساوياً لقيمة الانسجام $M_j$ لعقدة فئة الخرج زائد قيمة صغيرة موجبة $\varepsilon$ ($\varepsilon = M_j + \rho$). إذا انسجمت الفئة (وكان $\rho \le M_j$)، يحدث الطنين ويحدث معه تعليم ما. تعدل أوزان العقدة الرابحة **J** في فئة الخرج لتشمل جزءاً من التقاطع العائم لنموذج الدخل والأوزان الموجودة، وفقاً لمعادلة التحديث التالية:

$$\mathbf{W}_J^{new} = \alpha(\mathbf{I} \wedge \mathbf{W}_J^{new}) + (1 - \alpha)\,\mathbf{W}_J^{old}$$

حيث $0 < \alpha < 1$ هو عامل التعليم الذي يحدد مقدار التعليم. لاحظ أن القيم الكبرى لـــ $\alpha$ تعطي تعليماً أسرع لأن أوزان العقدة المختارة تكون متأثرة تأثراً كبيراً بنموذج الدخل. يستمر تدريب الشبكة في ممرات عديدة عبر مجموعة التدريب (أدوار عديدة) حتـــى تستقر الأوزان. بعدما إتمام عملية التدريب على مجموعة تمثيلية واستقرار الأوزان، تصبح

الشبكة جاهزة للعمل مع أي تصنيف لنماذج جديدة.

خلال العمل، يجري دوماً تخصيص فئة لنموذج دخل على أساس العقدة في فئة الخرج التـــي تربح المنافسة.

تستطيع شبكات ARTMAP العائمة تعلم مجموعات تجمع (مجموعات ضمن التجمع الواحد) أو تطبيقات عامة (mapping) من فراغ ذي بعد n إلى فراغ ذي بعد m، بأسلوب مشابه كثيراً للشبكات المتعددة الطبقات الأمامية التغذية المدروسة من قبل.

## 9.14 تطبيقات تستعمل شبكات نظرية الطنين المتكيف

## Applications using ART networks

سنصف في هذه الفقرة بعض التطبيقات النموذجية لشبكات نظرية الطنين المتكيف المستعملة لتنفيذ التجارب ولحل المسائل العملية الخاصة، بما في ذلك، تشخيص الخطأ وتحديد مكانه في عملية تصميم الدارات الرقمية، ودمج معطيات عدة حساسات رادار وملاحقة، وتحليل تجارة الأسهم والمبادلات المالية، ونظام استرداد المعطيات، وعمليات المراقبة، والتحكم. سنناقش ثلاثة من هذه التطبيقات فيما يلي.

### 1.9.14 تشخيص الخطأ وتحديد مكانه في عملية تصنيع الدرارات الرقمية

استعملت شبكات نظرية الطنين المتكيف لكشف وتحديد الأخطاء المتوضعة في الدارات الرقمية خلال عملية التصميم، والتركيب، وعمل الدارة. دربت الشبكة أولاً لتعرف الأخطاء وأماكن توضعها، من مجموعة تدريب أشعة الدخل التـــي تمثل نماذج خطأ معروفة ونماذج علامات عدم الخطأ (fault-free) المؤلفة من مداخل متسلسلة من نقاط التحكم وقياسات الاختبار عبر الدارة.

يمكن أن تولد معطيات التدريب بواسطة خوارزميات (Fugiwara وShimono عام 1983[28])، وبواسطة محاكيات لوح الدارة، أو يدوياً بتقديم الأخطاء المعروفة (حدوث قصر في الدارة). فور انتهاء تدريب الشبكة على عدد ضخم من الأزواج نوع خطأ/نموذج الدخل، يمكن أن تأخذ مكانها في عملية التصنيع لتقوم بتعلم جديد مستمر لأنواع خطأ جديدة غير معروفة قد تواجه عملية التصنيع.

تستطيع كل من شبكتــي نظرية الطنين الأولى وARTMAP العائمة تشخيص أخطاء الدارات الرقمية. في بعض الحالات، يمكن إنجاز تشخيص فعال بواسطة شبكتين تعتمدان نظرية الطنين المتكيف، الأولى تتعلم الأخطاء ونماذج عدم الخطأ، والأخرى تتعلم الأزواج المتعلقة بمكان ونوع الخطأ (Kalkunte عام 1992[240]).

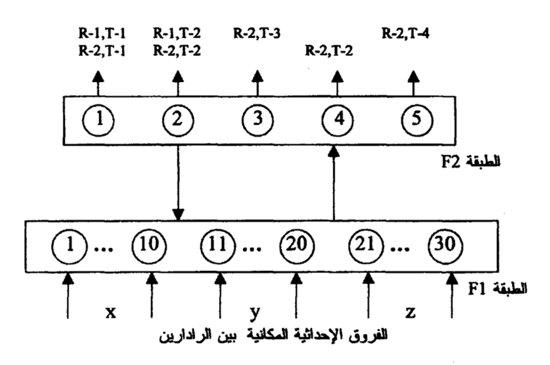
### 2.9.14 دمج معطيات عدة حساسات رادارية وملاحقة الهدف

وصفت عملية دمج معطيات عدة حساسات باختصار في فصول سابقة. في بعض التطبيقات، تقدم المعطيات من منابع متعددة إلى الشبكة بهدف تركيب المعطيات وإنجاز مهام التصنيف/التعرف والمراقبة والتحكم. المسألة الموصوفة هنا، تستعمل شبكة نظرية الطنين المتكيف الثانية لتركيب معطيات حركية (أماكن بإحداثيات ثلاثية x-y-z) متولدة من منبعين راداريين مختلفين. الراداران يلاحقان أهدافاً جوية متعددة في الزمن الحقيقي، وستدمج معطيات الهدف بواسطة الشبكة لتحديد أي من الأزواج هدف/رادار ينتمي إلى نفس المسار. بهذه الطريقة، يصبح تعقب الهدف واقتفاء أثره معقولاً أكثر، ودقيقاً. إضافة إلى ذلك، يمكن أن تصبح مهام دمج أخرى ملائمة أكثر مع مداخل منابع متعددة (أي تعريف الهدف، والوضع، وتخصيص المعالجة).

تتضمن شبكة نظرية الطنين المتكيف الثانية المستعملة في هذه المسألة 30 عقدة دخل في الطبقة F1 لقبول عشر عينات دخل، كل منها عبارة عن معطيات مكانية x-y-z من الرادارين. معطيات المكان المقدمة إلى الشبكة هي فعلياً الفرق المعياري في المكان x-y-z المأخوذ من الرادارين لعشر عينات خلال عشر ثوان .

توافق عقد الخرج فئات مسار/رادار. مثلاً، إذا لاحق الراداران مسار خمسة أهداف حيث الرادار -1 (R-1) يلاحق الأهداف T-1 وT-2 وT-5، والرادار -2 (R-2) يتعقب أثر الأهداف T-1 وT-2 وT-3 وT-4، فإن الشبكة ستصنف الزوج {(R-1 ,T-1) , (R-2 ,T-1)} كمسار مفرد (فئة خرج واحدة لشبكة نظرية الطنين المتكيف)، والزوج {(R-1,T-2) , (R-2,T-2)} كمسار مفرد آخر، وT-3 وT- 4 وT-5 كفئات منفصلة، ومن ثم سيكوت للشبكة خمس فئات خرج، كما هو موضح في (الشكل 20.14).

سيكون دخل الشبكة 30 رقماً حقيقياً توافق الفروق المعيارية بين معطيات التوضع الإحداثية الرادارية. تم تعديل وسيط الاحتراس باستعمال مسارات معروفة حتـــى أعطت الشبكة فئات صحيحة. بعد تمرين تدريب مختصر، صنفت الشبكة تصنيفاً صحيحاً كل مسارات الأهداف لمعطيات المحاكاة ومعطيات مسارات الأهداف الحقيقية. قورنت النتائج مع طرق الدمج الإحصائية التقليدية الأخرى.




**الشكل 20.14**: تصنيف ملاحقة رادارية من معطيات الفروق المكانية

ليكن الشعاعان $\mathbf{Y}_i(k)$ و$\mathbf{P}_i(k)$ يشير كل منهما إلى شعاع حالة المسار الفعلي رقم i ‎(i = 1 ,2)، ومصفوفة التباين المتبادل عند اللحظة k على الترتيب (R-1 وR-2)، وليكن $\mathbf{E}_{12} = \mathbf{Y}_1(k) - \mathbf{Y}_2(k)$ شعاع الفرق بين حالة المسار الفعلي لكلا الرادارين. أيضاً، ليكن $\hat{\mathbf{Y}}_i(k)$ هو شعاع حالة المسار المقدر و$\hat{\mathbf{E}}_{12} = \hat{\mathbf{Y}}_1(k) - \hat{\mathbf{Y}}_2(k)$ شعاع الفرق بين الحالة المقدرة لكلا الرادارين. ولنعرف $\acute{\mathbf{E}}_{12} = \mathbf{E}_{12}(k) - \hat{\mathbf{E}}_{12}(k)$ شعاع الفرق بين الحالة الفعلية والمقدرة لكلا الرادارين.

تعطى مصفوفة التباين المتبادل لـــ $\acute{\mathbf{E}}_{12}$ بالشكل التالي:

$$(\mathbf{P}_1(k) - \mathbf{P}_2(k) = \sum_{12}$$ (في حال كون الرادارين مستقلين

ستكون شروط الاختبار:

$$H0 : E_{12} = 0 \text{ (نفس الأهداف) }$$

$$H0 : E_{12} \neq 0 \text{ (أهداف مختلفة) }$$

باستعمال عتبة احتمال الارتباط الخاطئ $P_\alpha = 0.01$، استعمل توزيع كاي مربع $\chi^2$ الإحصائي (Chi-squared) (حيث لـ e توزيع كاي مربع) بـ 30 درجة حرية لاختبار H0. لمزيد من الإيضاح سنعرف باختصار هذا التوزيع. يعتبر المقدار $\chi^2$ مقياساً لمدى الاختلاف بين التكرارات المشاهدة والتكرارات المتوقعة. إذا كانت لدينا المتحولات العشوائية التالية:

$$X_1, X_2, \ldots, X_k$$

التـــي تعبر عن تحقق أحداث معينة عدد من المرات. وإذا كانت التكرارات المشاهدة هي:

$$\theta_1, \theta_2, \ldots, \theta_k$$

وكانت التكرارات المتوقعة للأحداث هي:

$$e_1, e_2, \ldots, e_k$$

فإن علاقة الارتباط التالية:

$$\chi^2 = \frac{(\theta_1 - e_1)^2}{e_1} + \frac{(\theta_2 - e_2)^2}{e_2} + \ldots + \frac{(\theta_k - e_k)^2}{e_k} = \sum_{i=1}^{k} \frac{(\theta_i - e_i)^2}{e_i}$$

تسمى بمدى المقياس للاختلاف بين التكرارات المشاهدة والتكرارات المتوقعة.

إن العدد k-1 يمثل عدد درجات الحرية التـــي يمكن أن تتمتع بها المتحولات $X_1, X_2, \ldots, X_k$، ذلك لأن المتحولات العشوائية يمكن أن ترتبط فيما بينها بالعلاقة:

$$X_1 + X_2 + \ldots + X_n = n$$

ولكي نحصل على واحدة منها فإننا سنفتقدها من المجموع. أخيراً، يمكن القول باختصار، إن $\chi^2$ لا يصح استعماله إلا في حالة $n \geq 50$ و $e_i > 5$، وإذا كانت قيمة $\chi^2$ معدومة فإن التكرارات المشاهدة تكون مساوية تماماً للتكرارات المتوقعة، وتعنـــي $\chi^2 > 0$ وجود تفاوت واضح بين التكرارات المشاهدة والمتوقعة، وكلما كانت قيمة $\chi^2$ كبيرة كان التفاوت كبيراً.

أما في حالة شبكة نظرية الطنين المتكيف، تعطي نتائج شروط الاختبار النموذجية أزواجاً

صحيحة للمسارات الرادارية التالية:

{(R-1 ,T-1) , (R-2 ,T-1)} كمسار مفرد

والزوج {(R-1 ,T-2) , (R-2 ,T-2)} كمسار مفرد آخر

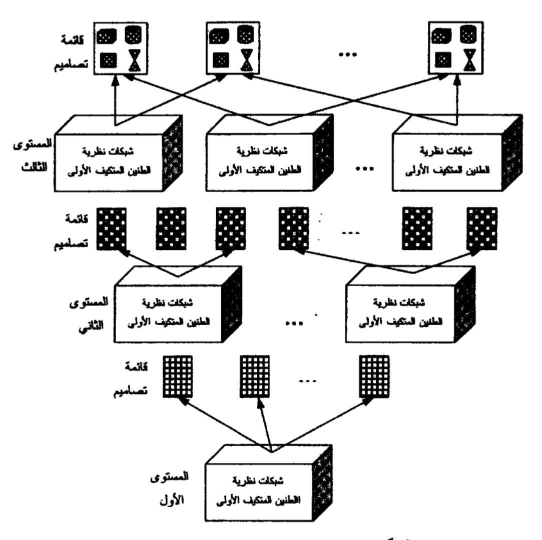
وT-3 وT-4 وT-5 كفئات منفصلة.

### 3.9.14 نظام استرداد المعطيات العصبوني

ترغب الشركات التـــي تصنع عدداً كبيراً من الأنظمة، مثل شركة طيران Boeing، بتوليد مئات أو حتـــى آلاف من تصاميم المكونات المختلفة. وبغياب خطة لفهرسة التصاميم، قد يعاد تصميم نفس المكونات إما بواسطة مجموعات تصميم مختلفة متوضعة بمواقع مختلفة وإما بواسطة نفس المجموعة بعد مرور بعض الوقت. سينتشر هذا الضياع غير المقصود للموارد في قطاعات التصنيع (Smith عام 1993[241])، ومن غير الضروري إضاعة الوقت والمال. تعتمد أنظمة الاسترداد التقليدية على فهرسة أو ترميز ملامح الجزء، وتعتبر هذه الأنظمة صعبة التحقيق والمتابعة. قادت هذه الاعتبارات قسم الخدمات الحاسوبية في شركة Boeing إلى إيجاد حل مبنـــي على شبكة عصبونية تجمع آلياً وتستعيد الأجزاء المصممة، اعتماداً على معطيات هندسية مستخرجة مباشرة من رسومات CAD (التصميم باستخدام الحاسوب).

استعملت شبكة نظرية الطنين المتكيف الأولى مبنية على نظام استرداد لتصنيف وتخزين التصاميم لقرابة 20000 جزء جوي مختلف. جرى تحقيق النظام بتسلسل تصاعدي من ثلاثة مستويات لوحدات قياسية من شبكات نظرية الطنين المتكيف، حيث يوافق كل مستوى مجموعة من الملامح.

استعمل المستوى الأول للشبكات لاختبار التصاميم المخزنة التـــي نظمت في مجموعات على أساس الشكل. اختير المستوى الثانـــي للشبكات على أساس الانحناءات في الأجزاء، واختيرت الطبقة الثالثة على أساس الثقوب في الأجزاء. تعطي هذه البنية المستخدم إمكانية التمييز على أساس الشكل فقط، أو على أساس الشكل والثقوب، أو على أساس الشكل والانحناءات، أو على أساس الشكل والثقوب والانحناءات.




**الشكل 21.14**: بنية نظام استرداد المعلومات العصبوني

نظام استعادة المعلومات العصبوني NIRS؛ (Neural Information Retrirval System) ككل موضح في (الشكل 21.14). يمثل كل صندوق في الشكل وحدة قياسية لشبكة نظرية الطنين المتكيف. تجمع الوحدات القياسية لشبكات نظرية الطنين المتكيف على شكل دارة كبيرة (المستوى). توافق كل دارة كبيرة ملمحاً وظيفياً. مثلاً، تختار الدارة الكبيرة للمستوى الأول على أساس شكل الأجزاء، ويختار المستوى الأعلى الثاني على أساس الثقوب، ويختار المستوى الأعلى الأخير على أساس الانحناءات. يمكن أيضاً اختيار تجميع الشكل والثقوب والانحناءات كمعيار للاختيار، ويجري توليد قوائم مناسبة من مواصفات الدخل للشبكات.

تسمح وسطاء الاحتراس للمستعمل بتغيير درجة الانسجام المختارة على كل الملامح المختارة. وهكذا، يمكن أن يسترد مجال التصاميم من عدد ضخم من الوحدات المتشابهة قليلاً

إلى مجموعة صغيرة من التصاميم المتشابهة كثيراً. عند حدوث شك، تضع أخفض الوحدات القياسية التصميم على أحد تجمعاتها. تمثل التجمعات عند هذا المستوى ملخصاً عاماً أكثر للتصاميم المخزنة. عند اختيار التجمع الرابع عند المستوى الأول، تصبح الوحدة القياسية في المستوى الأعلى التالي المرافقة لهذه المجموعة فعاله، وتضع هذه الوحدة القياسية التصميم على أحد تجمعاتها وتتكرر العملية.

دربت الشبكة بنفس الطريقة في المستويات الثلاثة. عند المستوى الأخفض تشكل معطيات التصميم على الحاسوب شعاعاً ثنائياً من مشهد ظليلي للجزء كصورة ثنائية البعد بـ 400 × 400 عنصر صورة. سيكون هذا الشعاع دخل الشبكة لتصنيف التجمعات. استعملت مداخل متشابهة في حالة صورة أماكن الثقوب والانحناءات في الجزء.

استغرق زمن التدريب تقريباً 12 ساعة وصرف معظم الوقت على توليد تمثيلات التصاميم باستخدام الحاسوب. يقع زمن الاسترداد باستخدام حاسوب شخصي بين 30 و45 ثانية.

## 10.14 تمارين

1-14 لتكن لدينا شبكة نظرية الطنين المتكيف الأولى بأربع وحدات تجمع في الطبقة F1 وثلاث وحدات في الطبقة F2. بعد تدريبها كانت الأوزان بالقيم التالية:

| أوزان أدنى لأعلى $v_{ij}$ | | | أوزان أعلى لأدنى $w_{ji}$ | | | |
|---|---|---|---|---|---|---|
| 0.67 | 0.0 | 0.2 | 1 | 0 | 0 | 0 |
| 0.0 | 0.0 | 0.2 | 0 | 0 | 0 | 1 |
| 0.0 | 0.0 | 0.2 | 1 | 1 | 1 | 1 |
| 0.0 | 0.67 | 0.2 | | | | |

حدد مصفوفات القيم الجديدة للأوزان بعد تقديم الشعاع (0, 0, 1, 1) في حالة:

1. وسيط الاحتراس يساوي 0.3
2. وسيط الاحتراس يساوي 0.7

2.14 ليكن لدينا شبكة نظرية الطنين المتكيف الأولى بتسع وحدات في الطبقة F1 ووحدتي تجمع في الطبقة F2. أصبحت قيم الأوزان بعد تدريب ما كما يلي:

أوزان أدنى لأعلى $v_{ij}$

| | |
|---|---|
| 1/10 | 1/3 |
| 1/10 | 0 |
| 1/10 | 1/3 |
| 1/10 | 0 |
| 1/10 | 1/3 |
| 1/10 | 0 |
| 1/10 | 1/3 |
| 1/10 | 0 |
| 1/10 | 1/3 |

أوزان أعلى ـــ لأدنى $w_{ji}$

1 0 1 0 1 0 1 0 1

1 1 1 1 1 1 1 1 1

بعد تقديم النموذج التالي إلى الشبكة (1, 1, 1, 1, 0, 1, 1, 1, 1). احسب عمل الشبكة إذا كان:

1. وسيط الاحتراس 0.5
2. وسيط الاحتراس 0.8

3.14 لتكن لدينا شبكة نظرية الطنين الثانية بوحدتي دخل (n = 2). أثبت أن استعمال وسـيط كبت الضجيج ذي القيمة 0.7 سـيجبر نماذج الدخـل التالية: (0.71, 0.69) و(0.69, 0.71) إلى تجمعات مختلفة. ما هو الدور الذي سيؤديه وسيط الاحتراس في هذه الحالة؟.

4.14 لتكن لدينا شبكة نظرية الطنين المتكيف الثانية المصممة لتجميع أشعة الدخل التالية:

(0.8, 0.6, 0.0) (0.6, 0.8, 0.0)

(1.0, 0.0, 0.0) (0.0, 1.0, 0.0)

$$(0.0, 0.6, 0.8) \qquad (0.0, 0.8, 0.6)$$

$$(0.0, 0.0, 1.0)$$

تحت أية ظروف ستجمع الشبكة أول شعاعي دخل هما (0.0 ,0.8 ,0. 6)، (0.0 ,0.6 ,0.8) معاً؟ ومتـــى ستجمع (0.0 ,0.8 ,0.6) مع (0.0 ,1.0 ,0.0)؟ استعمل وسيط تخميد ضجيج مساو إلى 0.577، واعتبر قيماً مختلفة لوسيط الاحتراس ولقيم الأوزان الأولية. اعتمد التعليم السريع.

5.14 تـــابع تنفيذ التمرين 4.14 السابق بافتراض أن الشبكة جمعت الأشعة (0.0 ,0.8 ,0.6) و(0.0 ,0.6 ,0.8) معاً. ماذا سيحدث إذا قدم الشعاع (0.0 ,0.84 ,0.55) إلى الشبكة؟. هل لقيم وسيط الاحتراس أو للقيم الأولية للأوزان لوحدات التجمع التـــي لم تتعلم أية نماذج تأثير في عمل الشبكة؟.

6.14 أثبت أن الشعاع **u** لن يتغير خلال التكرارات في الطبقة F1 في حالة النموذج الأول المتعلم بواسطة وحدة التجمع. إن قيم الأوزان المستقرة في حالة التعليم السريع في النموذج الأول المتوضع على التجمع ستوجد إذا نفذنا ما يلي في شبكة نظرية الطنين المتكيف الثانية:

$$\frac{dw_{Ji}}{dt} = du_i + d(d-1)w_{Ji}$$
$$= du_i + d(d-1)w_{Ji}$$
$$w_{Ji} = \frac{1}{1-d}u_i$$

وبأسلوب مشابه:

$$\frac{dv_{iJ}}{dt} = du_i + d(d-1)v_{iJ}$$
$$= \mathrm{d}u_i + d(d-1)v_{iJ}$$
$$v_{iJ} = \frac{1}{1-d}u_i$$

قيمة **u**، عندما تكون وحدة F2 الرابحة مختارة، هي ببساطة شعاع الدخل المعياري بأية مركبات أقل من وسيط تخميد الضجيج الموضوع بقيمة تساوي الصفر.

7.14 أثبت أن تكرارين للإشارة خلال الطبقة F1 كافيان لتخميد كل الضجيج. (يمكن أن

تعتبر أن تفعيلات بعض الوحدات تتغير بعد هذه اللحظة، لكن ليس تفعيلات الوحدات التـــي تعين وحدات F2 الرابحة أو التـــي تحدد تصفير أو قبول الوحدة الرابحة).

ابدأ بكل التفعيلات بقيمة الصفر، وافرض $e = 0$ أيضاً. افرض أن المركبة الأولى للشعاع **x** (شعاع الدخل يعد جعله معيارياً بطول واحدي) تسقط تحت قيمة وسيط تخميد الضجيج (على التكرار الأول) والمركبات الأخرى ليست كذلك. عرف الشعاع $\mathbf{ss} = (0, s_1, \ldots, s_n)$ كشعاع دخل، بمركبة مخمدة الضجيج بواسطة تابع التفعيل الموضوع بقيمة الصفر.

1. احسب تفعيلات الوحدات F1 من التكرار الأول. احسب التفعيلات **u** و **t** للتكرار الثاني.
2. أثبت أن

$$\|ss\| + a \leq \|t\| \leq \|I\| + a$$

3. باستعمال النتائج من الطلب 2 السابق، أثبت أن نظيم **t** يزداد من التكرار الأول إلى الثانـــي. عـــلـــى الـــتكـــرار الأول، $\mathbf{t} = (s_1, s_2, \ldots, s_n)$، وعـــلـــى التكـــرار الثاني، $\mathbf{t} = (s_1, s_2 + au_2, \ldots, s_n + au_n)$.
4. أثبت أن مركبات **x** التـــي كانت بقيمة الصفر على التكرار الأول ستأخذ القيمة صفر ثانيةً على التكرار الثاني، وأن المركبات التـــي لم تكن بقيمة الصفر في التكرار الأول لن تكون بقيمة الصفر في التكرارات اللاحقة.

8.14 باستعمال التعليم السريع، أثبت أن كبت الضجيج يمكن أن يساعد على منع عدم الاستقرار في عملية تجميع النماذج باعتبار إنجاز شبكة نظرية الطنين المتكيف على نماذج الدخل التالية:

النموذج 10 = (0.984798, 0.173648)

النموذج 20 = (0.939683, 0.342017)

النموذج 30 = (0.866018, 0.499993)

النموذج 40 = (0.766034, 0.642785)

النموذج 50 = (0.642785, 0.766034)

النموذج 60 = (0.499993, 0.866018)

النموذج 70 = (0.342017, 0.939683)

النموذج 80 = (0.173648 ,0.984798)

باستعمال قيم قياسية للوسطاء (a = 10, b = 10, c = 0.10, d = 0.90)، مع وسيط احتراس بقيمة 0.99 وأوزان أولية أدنى لأعلى تساوي (6.50, 6.50). استعمل حقيقة أنه في التعليم السريع، كل وحدة تجمع تتعلم نموذج الدخل الحالي تماماً.

قدم نماذج الدخل بالترتيب التالي: النموذج 40، النموذج 30، النموذج 20، النموذج 10، النموذج 40، النموذج 50، النموذج 60، النموذج 70، النموذج 80، النموذج 20، النموذج 30، النموذج 40.

1. استعمل وسيط تخميد ضجيج يساوي الصفر
2. استعمل وسيط تخميد ضجيج يساوي 0.2

9.14 اكتب برنامجاً لأداء عمل شبكة عصبونية صنعية بنظرية الطنين المتكيف الأولى. اكتشف إنجاز الشبكة لتراتيب دخل مختلفة لنماذج التدريب المستعملة في الأمثلة المشروحة ضمن الفصل.

10.14 اكتب برنامجاً لأداء عمل شبكة عصبونية صنعية بنظرية الطنين المتكيف الثانية، باستعمال التعليم السريع أو البطيء، وبالاعتماد على عدد أدوار تدريب وعدد تكرارات تحديث أوزان منفذة لكل تجربة تعليم. استعمل هذا البرنامج لاستكشاف العلاقات بين التعليم السريع والتعليم البطيء لنماذج دخل متنوعة.

11.14 بسبب أن شبكة نظرية الطنين المتكيف الثانية تجعل مداخلها معيارية، من المستحسن أحياناً إحداث مركبة إضافية لكل من الأشعة قبل تقديم المعطيات إلى الشبكة. هذه المركبة الإضافية ركبت بحيث يكون للأشعة الجديدة نفس المركبات الأولى كما في الأشعة الأصلية، لكن الأشعة الجديدة سيكون لها نفس النظيم (Dayhoff، عـــام 1990[188]). يمكن أن يختار N ليكون أي عدد أكبر من أكبر نظيم للأشعة الأصلية.

طبق هذه العملية على نماذج شجرة العبور. وباستعمال N = 10، تحصل على النماذج الموضحة في الجدول التالي. المركبة السادسة من كل شعاع هي الجذر التربيعي للكمية N ناقص نظيم الشعاع الأصلي.

باستعمال هذا الشكل من المعطيات، كرر مثال شجرة العبور. قارن وناقش نتائجك.

| النموذج | المركبات | | | | | |
|---|---|---|---|---|---|---|
| A | 1.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.9498 |
| B | 2.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.7979 |
| C | 3.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.5393 |
| D | 4.0 | 0.0 | 0.0 | 0.0 | 0.0 | 9.1615 |
| E | 5.0 | 0.0 | 0.0 | 0.0 | 0.0 | 8.6602 |
| F | 3.0 | 1.0 | 0.0 | 0.0 | 0.0 | 9.4868 |
| G | 3.0 | 2.0 | 0.0 | 0.0 | 0.0 | 9.3273 |
| H | 3.0 | 3.0 | 0.0 | 0.0 | 0.0 | 9.0553 |
| I | 3.0 | 4.0 | 0.0 | 0.0 | 0.0 | 8.6602 |
| J | 3.0 | 5.0 | 0.0 | 0.0 | 0.0 | 8.1240 |
| K | 3.0 | 3.0 | 1.0 | 0.0 | 0.0 | 9.0000 |
| L | 3.0 | 3.0 | 2.0 | 0.0 | 0.0 | 8.8317 |
| M | 3.0 | 3.0 | 3.0 | 0.0 | 0.0 | 8.5440 |
| N | 3.0 | 3.0 | 4.0 | 0.0 | 0.0 | 8.1240 |
| O | 3.0 | 3.0 | 5.0 | 0.0 | 0.0 | 7.5498 |
| P | 3.0 | 3.0 | 6.0 | 0.0 | 0.0 | 6.7823 |
| Q | 3.0 | 3.0 | 7.0 | 0.0 | 0.0 | 5.7445 |
| R | 3.0 | 3.0 | 8.0 | 0.0 | 0.0 | 4.2426 |
| S | 3.0 | 3.0 | 3.0 | 1.0 | 0.0 | 8.4852 |
| T | 3.0 | 3.0 | 3.0 | 2.0 | 0.0 | 8.3066 |
| U | 3.0 | 3.0 | 3.0 | 3.0 | 0.0 | 8.0000 |
| V | 3.0 | 3.0 | 3.0 | 4.0 | 0.0 | 7.5498 |
| W | 3.0 | 3.0 | 6.0 | 1.0 | 0.0 | 6.7087 |
| X | 3.0 | 3.0 | 6.0 | 2.0 | 0.0 | 6.4807 |
| Y | 3.0 | 3.0 | 6.0 | 3.0 | 0.0 | 6.0827 |
| Z | 3.0 | 3.0 | 6.0 | 4.0 | 0.0 | 5.4772 |
| 1 | 3.0 | 3.0 | 6.0 | 2.0 | 1.0 | 6.4031 |
| 2 | 3.0 | 3.0 | 6.0 | 2.0 | 2.0 | 6.1644 |
| 3 | 3.0 | 3.0 | 6.0 | 2.0 | 3.0 | 5.7445 |
| 5 | 3.0 | 3.0 | 6.0 | 2.0 | 5.0 | 4.1231 |
| 4 | 3.0 | 3.0 | 6.0 | 2.0 | 4.0 | 5.0990 |
| 6 | 3.0 | 3.0 | 6.0 | 2.0 | 6.0 | 2.4494 |

# الأنظمة العصبونية العائمة، الحساب المرن الخوارزميات الوراثية، شبكات المنطق العصبوني

# Neuro-Fuzzy Systems, Soft omputing,Genetic Algorithms Neuro-Logic Networks

في هذا الفصل، سنحيد قليلاً عن نموذج الفصول السابقة. فعوضاً عن اعتبار بنى الشبكات العصبونية الصنعية الخاصة أو تطبيقاتها، سينصب اهتمامنا على المواضيع التي لها علاقة قريبة من الشبكات العصبونية. سنبحث بوجه خاص في علاقة المنطق العائم مع الشبكات العصبونية الصنعية وتركيب الاثنين معاً لبناء الأنظمة العصبونية العائمة. وسنناقش أيضاً، مثالاً آخر للتكيف الذاتي، هو الخوارزميات الوراثية. وسننظر في الطرائق التي تطبق فيها الخوارزميات الوراثية لإنشاء الشبكات العصبونية الصنعية والأنظمة العائمة.

أخيراً، سنصف صنفاً مختلفاً تماماً من الشبكات العصبونية المعروفة بشبكات المنطق العصبوني، وشبكات المعالجات المنطقية المتوازية. وفي الفقرة الأخيرة من هذا الفصل، بل ومن هذا الكتاب، سننظر في التوجهات المستقبلية لتطور الشبكات العصبونية الصنعية.

## 1.15 تمهيد

في جميع الفصول السابقة، تركز اهتمامنا على بنى الشبكات العصبونية الصنعية التقليدية وتطبيقاتها. بدأنا في دراسة العصبون التمثيلي البسيط لـ McCulloch-Pitts ومن ثم تقدمنا من فهم العصبونات المتكيفة البسيطة مثل **perceptron** وADALINE وصولاً إلى أنظمة الذاكرة العصبونية المتعددة والشبكات المتعددة الطبقات الأمامية التغذية مثل MADALINE وشبكات MLFF الشائعة بخوارزمية الانتشار الخلفي.

ودرسنا التعديلات المنفذة على شبكات التغذية الأمامية مثل شبكة طاقة Coulomb المخفضة، وشبكة الارتباط المتتالي، وشبكة التراجع العامة، وneocognitron، و....وغيرها.

وكذلك درسنا بنـــى الشبكات المتنوعة بوصلات التغذية العكسية، والشبكات التـــي تنفذ الحسابات التكرارية على مداخلها، بما في ذلك، شبكات التكرار البسيطة بطبقة واحدة أو طبقات قرينة، وشبكات هوبفيلد، وشبكات نظرية الطنين المتكيف، وآلة بولتزمان، وشبكات التكرار العامة.

وقد درسنا خوارزميات تدريب متنوعة للتكيف بمعلم وبدون معلم. وخلال ما سبق من فصول شرحنا أمثلة عديدة وتطبيقات متنوعة لجميع هذه الشبكات العصبونية الصنعية.

في هذا الفصل سنعرض خلاصة ما جرى التوصل إليه.

سننطلق من نموذج سابق، وبدلاً من النظر في مواضيع قليلة مرتبطة بطريقة عامة أكثر للشبكات العصبونية، سننظر كيف ترتبط الشبكات العصبونية مع المنطق العائم وأنظمة الحساب المرنة، وكيف تشترك الشبكات العصبونية والخوارزميات الوراثية ببعض العموميات.

أخيراً، سنلخص صنف الشبكات التـــي تساعد على إنشاء الجسر الممتد عبر الهوة الواسعة بين الحساب الرمزي باستعمال المنطق التقليدي كالتمثيل، وطرق حساب الشبكات العصبونية لتنفيذ تعليل الحس السليم (التعليل بالمنطق الإنساني).

## 2.15 أنظمة الحساب المرن Soft Computig Systems

أحد الأهداف الرئيسية للباحثين في الذكاء الصنعي هو بناء أنظمة حاسوب تقلد إمكانية البشر في تفهم وتعليل مشاكل الحياة اليومية الحقيقية ومن ثم تحلها. ومن الغريب، أن تقريبات الذكاء الصنعي التقليدي في حل هذه المشاكل أسست على تطبيق تمثيلات المعرفة التقليدية كمنطق التنبؤ (ثنائي القيمة) والحسابات الرمزية. مع العلم أن هذا الشكل من الحساب القاسي أسس على الدقة والتعيين والصرامة (الإحكام). وهذا مغاير تماماً لتعليل الإنسان الحقيقي أو مايعرف بتعليل الحس السليم؛ التعليل المبنـــي على التقريب بدلاً من الدقة لطرق الحساب.

يكون التعليل البشري عادة مغلفاً بالغموض؛ فهناك حقائق غير كاملة مرفقة بعدد ضخم من غير المؤكدات (أو المتشابهات). إن وضع نموذج لهذا التعليل بدقة يتطلب خطة تمثيل تقوم بالتقاط المشاعر الحقيقية الطبيعية للعملية. هنا يستطيع المنطق العائم أن يؤدي دوراً هاماً، حيث إنه يوفر أساساً مرناً وطبيعياً لتمثيل معرفة غير معينة. حيث، يربط المنطق العائم اللغة مع الحساب (التعليل) على أساس المتغيرات اللغوية والمؤهلات. يمكن أن تعطى المتحولات اللغوية مثل "العمر" أو "الطول" كلمات كقيم مثل "شاب" أو "معتدل" والتـــي تكون قابلة للمكافأة في توابع العضوية العائمة. وبالمثل، تكون المؤهلات اللغوية مثل "بعض" أو "كثير" أو "أقل من النصف" قابلة للمكافأة كمجموعات جزئية عائمة للصف الحقيقي الذي يوافق القيم غير الدقيقة للكمية.

تحوَّل المتحولات والمؤهلات إلى توابع العضوية العائمة (توزيعات احتمالية) التـــي تَفترض لها قيماً في المجال [0, 1]. لنفترض أن U وV وW هي متحولات لغوية وX وY وZ مجموعات كلية للقيم الموافقة التـــي يمكن أن تكون لهذه المتحولات.

يعطى النوع القياسي لقاعدة نظام خبير عائم R كما يلي:

R: إذا (U يكون A) و(V يكون B) عندئذ (W يكون D)

R : If (U is A) and (V is B) then (W is D)

حيث A وB وD مجموعات جزئية عائمة لـــ X وY وZ على الترتيب. هذا النوع من العبارات يوضح المعرفة حيث تقع قيم المتحولات U وV وW في المجموعة العائمة A وB وD.

مثلاً، إذا كان U يوافق عُمْر البراء وVيوافق براعته وWمقدرته على الارتداء لمفرده، عندئذ تكون المجموعة الشاملة X هي مجموعة من القيم الصحيحة مـــن الصفر وحتـــى 10، وY مجموعة من القيم الصحيحة من 1 وحتـــى 10 (دليل البراعة)، وZ مجموعة من القيم الصحيحة من 1 وحتـــى 10 (درجات الارتداء). المعرفة بأن البراء طفل ببراعة معينة ومن ثم تكون مقدرته على الارتداء بقدر محدود موضحة بواسطة رابط الافتـــراض "(U يكون A) و(V يكون B)" مع نتيجة مقدرته على الارتداء موضحة في التضمين "(W يكون D)". يتألف أساس قاعدة الخبرة الكاملة من عدة قواعد:

$R_1$ : إذا ($U$ يكون $A_1$) و($V$ يكون $B_1$) عندئذ ($W$ يكون $D_1$)

$R_2$: إذا ($U$ يكون $A_2$) و($V$ يكون $B_2$) عندئذ ($W$ يكون $D_2$)

... ... ...

... ... ...

$R_k$ : إذا ($U$ يكون $A_k$) و($V$ يكون $B_k$) عندئذ ($W$ يكون $D_k$)

ومع أن المنطق العائم يوفر خطاً قريباً بين اللغة الطبيعية و"تقريب التعليل الحسابـــي"فإن طرق الحساب العائمة لا تشمل المقدرة على التعلم المتكيف لإنجاز أعمال الذاكرة المترافقة ومستويات عالية من التسامح مع الضجيج وتشوهات النموذج (التـــي امتازت بها شبكاتنا العصبونية)، هذه المقدرات الضرورية لعدة مهام كالإدراك والتعليم والاستجابة للسلوك التنبئي. ولكن تستطيع الشبكات العصبونية الصنعية أن تؤدي بمهارة مثل هذه الأدوار. إذاً، بوضع كلا المقدرتين معاً، سيكون لدينا نظام حساب عصبونـــي عائم قوي. وهذا ما دعاه لطفي زاده عام (واضع أسس المنطق العائم) عام 1994 بالحساب المرن [132].

إن طرائق الحساب الناعم تجعل من الممكن وصف مفهوم صعب أو إيقاف سيارة بسهولة. فالوصف الكافي للمفهوم هو تقريب للواحد، ومكان واتجاه السيارة التـــي أوقفت لا يحتاج إلى أن يكون مثبتاً بدقة.

تساعد معالجات هذين الحقلين اللذين يتمم أحدهما الآخر على إيجاد أنظمة ذكية باعتبار أنهما شكلان لنموذج بشري للتمثيل والتعليل بدقة أكثر من التقريبات القاسية التقليدية. لذلك، فإن من الطبيعي أن يكون هذان الحقلان مستعملين معاً، ومركبين معاً في نظامٍ عصبونـــي عائم.

هناك اقتراحات عديدة نفِّذت في الأنظمة العصبونية العائمة بما في ذلك أعمال Teow وزملائه عام 1993 [242] وHsu وزملائه عام 1989 [243] وYager وزملائه [147] عام 1994.

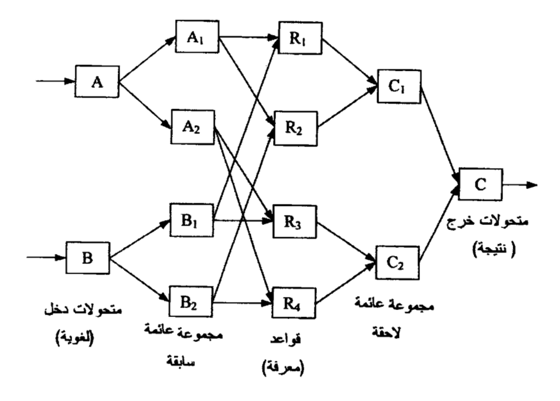
تختلف الطرائق المقترحة بكلا التمثيلين وبنـــى العمل.

سنلخص تقريباً واحداً تمهيدياً يكون الأساس في خصوصيات التصميم للهيكل العصبونـــي العائم.

في الحقيقة، لقد طورت الهياكل العصبونية العائمة التامة في معهد علم الأنظمة (Institute of Systems Science) التابع للجامعة الوطنية بسينغافورة (National University of Singapore) في بداية 1991.

الهياكل العصبونية العائمة هي أنظمة عامة تسهل إيجاد نظام عصبوني عائم لمسألة خاصة مطروحة للمعالجة. يلزم أن تكون فقط معرفة مرمزة ومضافة إلى النظام.

نظامنا العصبوني العائم الذي نحن بصدده مؤلف من خمس طبقات من العصبونات مع وصلات داخلية مختارة بتغذية أمامية، كما هو موضح في الشكل (1.15).




**الشكل 1.15:** شبكة عصبونية عائمة بسيطة

عصبونات الطبقة الأولى، العقد A وB، هي بقيم متحولات دخل لغوية موافقة لروابط في سوابق القواعد العائمة. كل من هذه الوحدات وصل إلى بعض الوحدات في الطبقة المخفية الأولى. قسمت عصبونات الطبقة المخفية الأولى إلى مجموعات جزئية، يوافق كل منها حقلاً عائماً. وصلت الوحدات من هذه المجموعات الجزئية إلى وحدات الطبقة المخفية الثانية؛ عصبونات القاعدة العائمة. توافق كل وصلة دخل إلى القاعدة رابطاً في سالف القاعدة.

مخارج وحدات القاعدة هي لواحق (عواقب) القاعدة. وصلت إلى عصبونات الطبقة المخفية الثالثة في مجموعات الحقل العائم بأسلوب مشابه لوحدات الطبقة المخفية الأولى. ستكون الطبقة الأخيرة النتيجة؛ وهي قيمة متحول لخرج العائم.

هناك أربع قواعد عائمة في الشكل (1.15) يمكن أن تفسر كما يلي (تابع اتجاه الأسهم في الشكل مع القاعدة):

$R_1$: إذا (A يكون $A_1$) و(B يكون $B_1$) عندئذ (C يكون $C_1$)

$R_2$: إذا (A يكون $A_1$) و(B يكون $B_2$) عندئذ (C يكون $C_1$)

$R_3$: إذا (A يكون $A_2$) و(B يكون $B_1$) عندئذ (C يكون $C_2$)

$R_4$: إذا (A يكون $A_2$) و (B يكون $B_2$) عندئذ (C يكون $C_2$)

طريقة الاستدلال العائمة المستعملة لهذا النظام هي استدلال تدفق قيمة الحقيقة TVFI (Truth Value Flow Inference)كما وصفت في الفصل الثالث ومن قبل Wang عام 1993[112]. تستعمل هذه الطريقة خطوة تعويم مبسطة عند خرج الشبكة دون الحاجة إلى إنجاز خطوة المحصلة الكلية. إنها تكافئ الطريقة التـــي اقترحها Mamdani عام 1974[244] (المركز المتوسط للمجموعات المركبة) وهي أكثر تعقيداً في بعض الحالات.

المداخل إلى الشبكة هي مجموعات عائمة معيارية محدبة، A ؛ أي لها على الأقل نقطة واحدة $x$ بقيمة انتماء $\mu_A(x)=1$ وفي مجال ما [1, r] الذي تكون فيه $\mu_A(x)=1$، نعرف نقطة المنتصف $P_A=(1+r)/2$ كنقطة تمثيلية للمجموعة العائمة A.

مررت المداخل المعيارية إلى وحدات مجموعة السوابق المختلفة عبر وصلات الدخل إلى وحدات السوابق. يوافق خرج وحدات السوابق قيم الانتماء لمجموعة الدخل العائمة. كل وحدة قاعدة لها قيم وزن تساوي $1/k$، حيث $k$ عدد المداخل إلى القاعدة. بهذه الأوزان، تنجز الشبكة عملية MIN (كما هو موصوف فيما بعد)، وهي عملية قياسية للتقاطع العائم. وبالمثل، لكل وحدة مجموعة لاحقة قيم أوزان تساوي الواحد تؤدي إلى حساب العملية MAX، باستعمال تابع التفعيل المعرف فيما يلي، وهي العملية القياسية لمجموعة الاجتماع. بعدئذ، تحسب وحدات الخرج النتيجة غير العائمة النهائية وفقاً للعلاقة التالية:

$$out = \frac{\sum_i p_i z_i}{\sum_i z_i}$$

حيث $z_i$ قيمة تفعيل المجموعة اللاحقة رقم i الواصلة إلى وحدة الخرج، و $p_i$ نقطة تمثيلية كما هو مذكور من قبل. تعطى الخوارزمية الكاملة للشبكة العصبونية العائمة كما يلي:

1. تفرز المداخل $x_i$ بترتيب متصاعد بإعطاء $x'_i \le x'_{i+1}$ وحساب الفروق $b_u = x'_{i+1} - x'_i$ مع $x'_0 = 0$.
2. حساب الأوزان المرافقة للمداخل $b_i$ للحصول على $\{w_1, w_2, \cdots, w_n\}$
3. حساب قيمة تفعيل الخرج:

$$output = \sum_{i=1}^{n} b_i F\left(\sum_{j=1}^{n} w_j\right)$$

حيث

$$F(x) = \begin{cases} 0 & x \le (1-1/n) \\ n(x+1/n-1) & (1-1/n) < x < 1 \\ 1 & x \ge 1 \end{cases} \qquad (1.15)$$

لقد استعمل في تدريب الشبكة نوع من تدرج هبوط الانتشار الخلفي للخطأ للتناغم الدقيق لطبقتـــي الأوزان (وضعت مبدئياً 1/k و 1 على الترتيب) لتنعيم عمليات MIN و MAX.

بنيت معادلات الإرسال الخلفي على قاعدة دلتا المعممة الموصوفة في الفصل السادس، لكن مع حد اشتقاق تجريبـــي (تنقيبـــي) (heuristic) مبنـــي على معادلة التفعيل (1.15). نفذت محاكيات الشبكة العصبونية العائمة لتقييم إنجازها في تعلم نسخة عائمة لقانون الغاز النموذجي PV= nRT، حيث P الضغط، وV الحجم، وT درجة الحرارة، وn وR ثوابت كيفية. بنيت الشبكة باستعمال أداة تدعى Flexi-NET طورها معهد علم الأنظمة (Institute of Systems Science). في هذا النظام، تم تعلم 18 قاعدة. تعطى قاعدة نموذجية متعلمة كمايلي:

إذا كان **الحجم** كبيراً جداً و**الضغط** عالياً جداً فإن **درجة الحرارة** ستكون T9. الكميات Tk (T9) في نتيجة القاعدة هي مجموعة لاحقة لدرجة الحرارة.

دربت الشبكة على 1000 مثال لعشرة أدوار. ولزم عشرة أدوار فقط لتناغم دقيق للأوزان لأن أمثلة التدريب كانت من قواعد معطاة بواسطة "خبراء".

للشبكات من هذا النوع محاسن عديدة بالمقارنة مع الأنظمة التقليدية، فهي قادرة مثلاً على ترميز معرفة خبرة معطاة في عبارات لغوية غير دقيقة، وحتـــى تحسين دقة المعرفة بواسطة أمثلة التدريب.و أكثر من ذلك، ليست الشبكة صندوقاً أسود، فالعمليات يمكن أن تكون مفهومة من خلال القواعد التـــي ترمزها.

استعملت الشبكات العصبونية والأنظمة الخبيرة عائمة المنطق أيضاً في التتميم عوضاً عن الطريقة التركيبية. مثلاً، استعملت الشبكات العصبونية لتسريع تصميم الأنظمة العائمة وتحسين إنجازها من خلال الدقة العالية في:

1. تحديد العدد الأولي للقواعد العائمة للنظام،
2. تحديد أفضل توابع الانتماء،
3. تعديل توابع الانتماء بأسلوب متكيف عندما يحدث تغير في الوسط المحيط.

يمكن أن تستعمل الشبكات العصبونية لتعديل نتائج التعليل العائم للنظام الخبير العائم. باختصار، يمكن أن تطبق الشبكات العصبونية والمنطق العائم بآن واحد ضمن نفس النظام لإنجاز مختلف، لكن لأغراض متتامة بعضها مع بعض أو منفصلة كأدوات لتحسين إنجاز النظام.

فهما متتامان فيما يلي:

+ يستطيع المنطق المبهم التعبير جيداً عن القيم النوعية للمنطق الإنسانـــي ويوفر أفعالاً مرنةً من خلال توابع الانتماء المستمرة (جيدة من أجل تطبيقات لتحكم وتطبيقات أخرى).
+ تعبر قواعد المنطق العائم عن مجال واسع من علاقات شرط/فعل، وبذلك تتطلب قواعد أقل من الأنظمة الخبيرة المبنية على المنطق التقليدي.
+ يمكن أن تتعلم الشبكات العصبونية صياغة توابع غير خطية معقدة من نماذج التدريب مثل أسطح توابع الانتماء المتعددة الأبعاد التـــي تكون صعبة التصميم (مثل درجة الحرارة، والرطوبة، وسرعة الريح).

+ يمكن أن تتعلم الشبكات العصبونية مهام متنوعة من أمثلة التدريب(الطنين المتكيف) بما في ذلك متتاليات النماذج الزمانية.

لقد وصلت إلى الأسواق أنظمة تستعمل المنطق العائم منذ سنين عديدة. وتستعمل الشركات اليابانية المنطق المبهم منذ 1980 في مختلف المنتجات الاستهلاكية، بالإضافة إلى عدد ضخم من الأنظمة الصناعية بما في ذلك عناصر التحكم الذكية (البارعة) في قطارات الأنفاق.

بدأت تطبيقات الأنظمة العصبونية العائمة تظهر أيضاً في الأسواق.

في 6 كانون الثاني (يناير) عام 1995 نشرت جريدة النجمة الماليزية، صفحة إعلان كاملة لثلاجة غولد ستار (Goldstar) التي تتحكم العصبونات العائمة فيها. من بين الأشياء الأخرى المعلن عنها، (وهناك الكثير من الأشياء التي مازالت داخل مخابر البحث قيد الكتمان)، كان ثلاجة ضخمة بستة أبواب استخدمت نظاماً ذكياً يشبه الدماغ البشري وكانت قادرة على تعلم وتخزين تصرفات المستعمل، ومكيفة لتحتفظ بدرجة حرارة داخلية ثابتة في كل وقت، وكان ذلك يعتبر البداية لقائمة طويلة من تطبيقات الأنظمة العصبونية العائمة المتطورة.

## 3.15 الخوارزميات الوراثية Genetic Algorithms (GAs)

اقترحت الخوارزميات الوراثية ودرست من قبل John Holland وزملائه عام 1975 [245] وذلك في جامعة Michigan. فقد بحث في الخوارزميات التي بنيت على ميكانيكيات الانتقاء الطبيعي والمورثات الطبيعية. وأنجز بحثاً متوازياً عشوائياً شاملاً للحل الأمثلي باستعمال حسابات بسيطة.

البداية مع مجتمع أولي للبنى الوراثية، وعمليات الإرث الوراثية المبنية على الانتقاء (Selection)، والتزاوج (mating)، والطفرة الوراثية (mutation) المنجزة لتوليد نسل (offspring) يتنافس على البقاء (البقاء للأصلح) لتركيب الجيل الثاني من بنى المجتمع.

تميزت الخوارزميات الوراثية بالانجاز الموثوق باستعمال أسرار التكيف والبقاء كما

نمذجت بعد تطور البيولوجيا. وبرهن نظرياً وتجريبياً أنها توفر بحثاً موثوقاً في المجالات المعقدة لكونها غير محددة بفرضيات صارمة للاستمرارية، ووجود الاشتقاقات، والشروط المقيدة الوحيدة النمط وغير ذلك.

لقد وجدت تطبيقات واسعة في مجالات الأعمال، والعلوم، والهندسة بما في ذلك، تعرف الأشكال، ووظائف الاستمثال، والجدولة، وتعليم الآلة، والتجميع، والتصميم الهندسي، وتصميم النظام الخبير، وعمليات التحكم، وتطبيقات أخرى عديدة.

في منهج الخوارزمية الوراثية، المعرفة ممثلة كحوض مجتمع $\prod(t)$ من M بنية وراثية عند اللحظة t:

| | | | | |
|---|---|---|---|---|
| $A_1$ | $A_2$ | $A_3$ | .... | $A_n$ |
| $B_1$ | $B_2$ | $B_3$ | .... | $B_n$ |
| $C_1$ | $C_2$ | $C_3$ | .... | $C_n$ |
| | | ............... | | |
| | | ............... | | |
| $M_1$ | $M_2$ | $M_3$ | .... | $M_n$ |

حيث [M] ,...., [B], [A] أنواع وراثية كل منها ممثل بواسطة N صبغياً متضادة في الصفات $X_1,X_2,X_3,\ldots,X_n$.

يمكن أن تُمَثِّل البنية الوراثية:

1. وسطاء توابع الانتماء العائمة
2. قواعد "إذا كان ....فإن...."
3. وصلات أو قيم الأوزان في الشبكة العصبونية
4. طح الاستجابة لنظام التحكم غير الخطي
5. حركات في لعبة الشطرنج
6. قيمة تابع

وهكذا. فإن البنـــى المختارة عموماً، هي سلاسل حرفية مثل سلاسل الخانات الثنائية أو الأرقام الحقيقية. وتختار العمليات اللازمة لنمذجة المورثات العصبونية، عموماً كتقاطع

(crossover)، أو كطفرة (mutation)، أو كقلب. (inversion) وقد وُصفت ببساطة في المثال التالي:

**التقاطع: "تزاوج بين سلسلتين حرفيتين"**

بعد التحويل | قبل التحويل

نقطة التقاطع تختار عشوائياً وهي هنا

$A_1\ A_2\ A_3\ A_4\ A_5\ A_6\ A_7 \Rightarrow A_1\ A_2\ A_3\ B_4\ B_5\ B_6\ B_7$

(أب النوع الوراثي 1) (النسل 1)

$B_1\ B_2\ B_3\ B_4\ B_5\ B_6\ B_7 \Rightarrow B_1\ B_2\ B_3\ A_4\ A_5\ A_6\ A_7$

(أب النوع الوراثي 2) (النسل 2)

تطبق عملية التقاطع على بنيتــي أب (أصل) مختارتين احتمالياً من المجتمع. بعدئذ، تختار نقطة التقاطع عشوائياً كجزء بين اثنين من صبغيات الأب (خانات أو أحرف)، والصبغيات التالية لنقطة التقاطع في الأب الأول تكون موضوعة بالتسلسل مع الصبغيات السابقة لنقطة التقاطع في الأب الثانــي والعكس بالعكس. السلسلتان الناتجتان من التسلسلات المتبادلة تكونان جيل الأبناء .

الطفرة هي عملية تغيير شكل أو حرف (charactar) واحد (بت واحدة) بسيطة على عضو مجتمع مفرد. يختار العضو باحتمال صغير، ويختار مكان الحرف (مكان الخانة) عشوائياً. بعدئذ يغير الحرف في ذلك المكان.

في حالة كون سلسلة الأب ثنائية، يؤخذ متمم الخانة المختارة (الصفر يصبح واحداً والواحد يصبح صفراً). طبقت هذه العملية بتكرار أقل من عمليات أخرى، نموذجياً أقل من 1% من الزمن. من المفيد إلغاء ممرات البحث غير المجدية أو عندما يصبح أعضاء المجتمع راكدين.

عملية الطفرة موضحة فيما يلي حيث عضو المجتمع افترض ليكون سلسلة ثنائية بسبع خانات:

**الطفرة: مشغل الطفرة العشوائي**

خانة الطفرة المختارة

| | |
|---|---|
| قبل الطفرة | $A_1\ A_2\ A_3\ A_4A_5\ A_6\ A_7$ |
| بعد الطفرة | $A_1\ A_2\ A_3\ A_4A_5\ A_6\ A_7$ |

عملية القلب أيضاً، هي عملية أب وحيد. كما في الحالة السابقة، يختار الأب على أساس توزع احتمالي، ويختار مكان الحرف في الأب عشوائياً. شكلت بنية النسل الناتجة بتسلسل الأحرف بعد نقطة القلب متبوعة بالأحرف السابقة لنقطة القلب.

عملية القلب موضحة كما يلي :

**القلب**

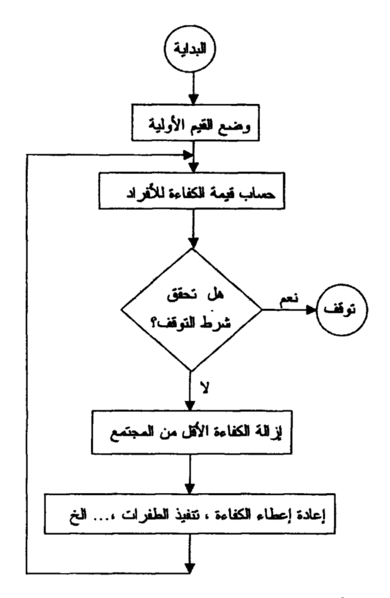
خانة الطفرة المختارة

| | |
|---|---|
| قبل الطفرة | $A_1, A_2, A_3, A_4, A_5, A_6, A_7$ |
| بعد الطفرة | $A_4, A_5, A_6, A_7, A_1, A_2, A_3$ |

بهذه العمليات الأساسية (ومن الممكن تنفيذ عمليات أخرى لم تذكر)، نفذت عملية بحث الخوارزمية الوراثية وفقاً للمخطط الصندوقي الموصوف في الشكل (2.15).

للبدء بحلقة البحث، يجب أن تكون الكفاءة أو قياس الإنجاز معرّفة كتابع لأعضاء المجتمع. ويجب أن تحدد القيم الأولية لوسطاء عديدة، بما في ذلك القيم الدنيا لحجم المجتمع، وتكرار الطفرة، والتقاطع. بعدئذ، يولد المجتمع البدائي (عشوائياً، من المحتمل مع بعض الشروط المقيدة)، وتحسب الكفاءة لكل فرد. إذا حققت كفاءة كل فرد معيار التوقف، ينتهي البحث وإلا يتولد جيل جديد بواسطة إعادة الإنتاج، والتحويل، والطفرة و/أو عمليات أخرى ممكنة.

أحد أسباب نجاح الخوارزميات الوراثية هو قدرتها على استغلال أفضل صناديق البناء للأفراد. فهي تميل إلى توليد إنجاز عال، خلال زمن قصير، وخطط بمرتبة أقل عندما يزداد عدد الأجيال المتعاقبة أسيّاً.

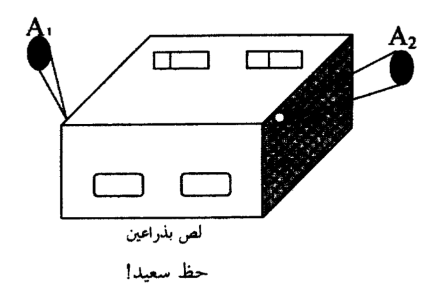



**الشكل 2.15**: مخطط صندوقي للخوارزمية الوراثية الأساسية

مثلاً، تعمل خطة ثنائية بـ k-bit كبنية انسجام نموذج لأي k bit، حيث يمكن أن تكون كل قيمة مكان خانة 0، أو 1، أو $ (مهما تكن). خطة الخانات الخمس التالية $0$10 تنسجم مع أربع سلاسل: 00010 و10010 و00110 و10110.

لقد ثبت أن عمل الخوارزميات الوراثية يشبه قضية اللص الذي له k ذراعاً (k-armed bandit problem) فهي تعطي أعداداً متزايدة أسياً من التجارب لمراقبة أفضل لـ k ذراعاً (خطط) في البحث عن حل أصغري للضياع المتوقع (Goldberg عام 1989 [194]).

سنصف باختصار مسألة اللص بذراعين (k = 2) الموضحة في الشكل (3.14).




**الشكل 3.15**: مثال لص ثنائي الذراع لعملية الخوارزمية الوراثية

من المعروف أن اللص الذكي هو الذي يتقن حركة ذراعه أو ذراعيه معاً بدقة وبسرعة متناهية لتناول النقود من أحد الجيوب الممتلئة دون أن يشعر المسروق منه ما فعل السارق. هذه قضية تقدير أمثلي لمقدرة كل ذرع أو الذراعين معاً بأقل ضياع ممكن في اختيار اليد المناسبة، والحركة المناسبة، والسرعة المناسبة، و...الخ.

الآن، نحتاج إلى تصميم آلة ذات ذكاء صنعي (كذكاء ذاك اللص) لفعل الأخاديد والشقوق (آلة تشقيب) بذراعين منفصلين لهما احتمالات دفع غير معروفة. من المعروف أن احتمالات الذراعين مختلفة، لكن الذراع ذا الدفع الأعلى غير معروف. بالطبع يرغب المرء بتحريك ذراع الدفع الأعلى فقط وبأسرع ما أمكن. مسألة اللص بـــ k ذراعاً هي تعميم مباشر لمسألة اللص الثنائي الذراع.

في الخوارزميات الوراثية، فإن مسألة اللص بـــ k ذراعاً التشابهية يمكن أن تستعمل لإثبات أن الخطط تنمو وفق العلاقة التالية:

$$m(H,t+1) = m(H,t)\frac{f(H)}{\bar{f}}$$

حيث m(H,t) عدد الأمثلة لنوع معطى H من الخطط عند اللحظة t، وf(H) متوسط الكفاءة للسلاسل الممثلة للخطط H، و $\bar{f}$ متوسط الكفاءة لمجتمع السلاسل بالكامل. إذا بقيت الخطط H أعلى المتوسط بواسطة $c\bar{f}$ (c > 0) عندئذ يعطىعدد الخطط H في اللحظة t بالعلاقة التالية:

$$m(H,t) = m(H,0)(1+c)^t$$

في مسألة اللص الثنائي الذراع، ليكن :

الذراع الأول $A_1$ له مكافئة: $r_1$ مع تباين $\sigma_1$

الذراع الثانـــي $A_2$ له مكافئة : $r_2$ مع تباين $\sigma_2$

حيث $r_1 \geq r_2$ وكلا $r_1$ و$r_2$ غير معروف. بعدئذ، يمكن إثبات أن الضياع المتوقع L يعطى بالعلاقة:

$$L(N,n) = |r_1 - r_2|[(N-n)q + (1-q)n]$$

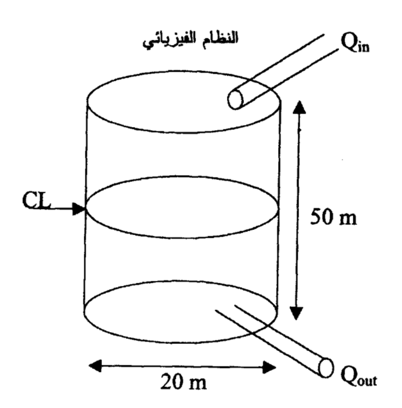
حيث N العدد الكلي للتجارب (محاولة)، وn عدد المحاولات لكل ذراع، وq احتمال الخطأ بعد 2n محاولة. من هذا، يمكن إيجاد المحاولات الأمثلية *n من العلاقة:

$$N - n^* \cong N \cong (8\Pi b^4 \ln N^2)^{1/2} \exp(n^*/2b^2)$$

حيث $b = \sigma_1/(r_1 - r_2)$.

لتوضيح بعض المفاهيم السابقة، سننهي هذه الفقرة بمثال لتطبيق الخوارزمية الوراثية في تصميم نظام المنطق العصبوني. يوضح المثال كيف ينفذ نظام عنصر التحكم بالمنطق العائم المصمم بخوارزمية وراثية بأسلوب أفضل مقارنة مع النظام التقليدي، نظام العمل بالتجربة والخطأ (Karr عام 1991[200]). يتألف النظام من خزان أسطوانـــي الشكل يحوي سائلاً كيميائياً معيناً يستجر من الخزان بعملية التصنيع.

المسألة هي التحكم في مستوى السائل بحيث أن الذي يوضع عند المستوى h في البداية يحافظ عليه مع انحراف صغير.




**الشكل 4.15**: نظام تزويد سائل متحكم فيه بالمنطق العائم

جرى تحقيق التحكم باستعمال عنصر تحكم في التدفق يعمل بالمنطق العائم. النظام موضح في الشكل (4.15). صممت توابع الانتماء لنظام التحكم بالاستعانة بالمهندس الخبير وباستعمال الخوارزمية الوراثية. وجرت مقارنة الإنجاز لتقريبين مختلفين. ولتوضيح أفضل لتقريب الخوارزميات الوراثية لمسألة كهذه، سنصف تفاصيل عملية التصميم الأمثلية.

يوصف النظام الفيزيائي بالمعادلة التالية:

$$h^{t+1} = h^{t} + \left[\frac{Q_{in} - Q_{out}}{A^{tank}}\right]\Delta t$$

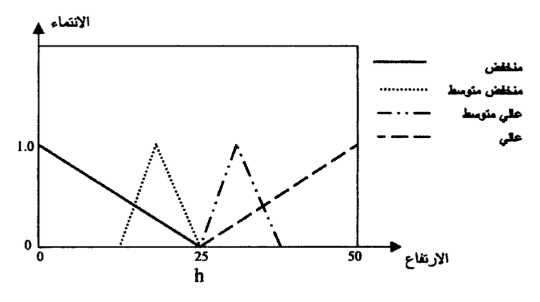
حيث h ارتفاع مستوى السائل، و$A_{tank}$ مساحة المقطع العمودي على محور الخزان، و $\Delta t$ خطوة الزيادة الزمنية، و$Q_{in}$, $Q_{out}$ مقدار التدفق الحجمي إلى داخل وخارج الخزان على الترتيب.

سمح لمعدلات التدفق لداخل وخارج الخزان أن تتغير من 0 إلى 200 متر مكعب كل ثانية مع خطوة زمنية تساوي ثانية واحدة.

متحولات القرار المستعملة بواسطة عنصر التحكم هي:

1. h ( أربع مجموعات عائمة: عال، عال متوسط، منخفض متوسط، منخفض)
2. dh/dt المعدل الزمنـــي لتغير ارتفاع السائل (خمس مجموعات عائمة: موجب كبير، موجب صغير، قرب الصفر، سالب صغير، سالب كبير)
3. تدفق السائل إلى داخل الخزان Qin
4. تدفق السائل إلى خارج الخزان Qout .

توابع الانتماء المستعملة لمتحولات القرار الأربعة هي توابع مثلثية قياسية، كما هو موضح في الشكل (5.15).



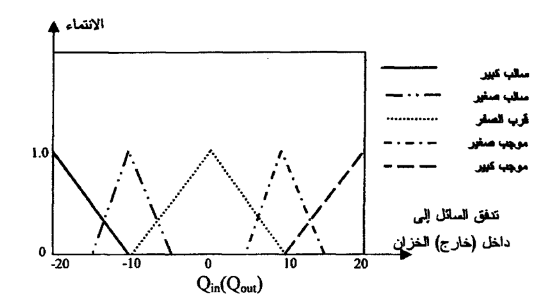

**الشكل (5.15)**(أ): تابع الانتماء لارتفاع مستوى السائل

التابع الموضوعي f المستعمل لتحديد نسبة الكفاءة للأفراد هو تابع الفروق المربعة في المستوى عن منتصف الخط (25 m):

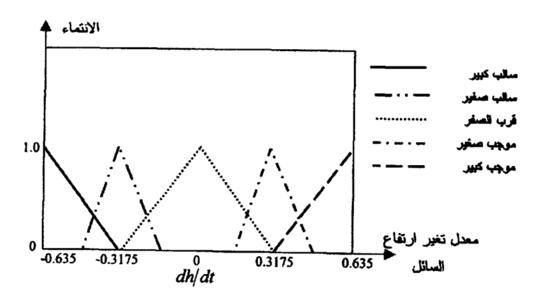
$$f(h,t)=\sum_{i=case1}^{case4}\sum_{j=0s}^{20s}\left(25-h_{ij}\right)^2$$

لاحظ حدود المجاميع في التابع الموضوعي.

جرى تضمين الدليل الزمنـــي لمعاقبة النظام في حالة الاستجابة البطيئة والحالات الأربع للشروط البدائية.اختيرت الحالات للتأكيد أن النظام يستطيع رفع وخفض مستوى السائل بفعالية متساوية.




**الشكل(5.15)(ب)** تابع الانتماء لتدفق السائل إلى داخل(خارج) الخزان




**الشكل (15.5)** (ج) تابع الانتماء لمعدل تغير ارتفاع السائل

المستويات المختارة هي:

| dh/dt | h | العدد |
|---|---|---|
| 0.6366 + | 00.0 | 1 |
| 0.6366 – | 50.0 | 2 |
| 0.3183 – | 10.0 | 3 |
| 0.3183 – | 40.0 | 4 |

ثمة وسطاء أخرى استعملت في تصميم الخوارزمية الوراثية مثل:

احتمال التحويل: 0.8

احتمال الطفرة: 0.01

حجم المجتمع: 500

التوليدات الأعظمية: 80

طول السلسلة: 132 خانة (خانات لكل وسيط)

يتألف المجتمع من 500 سلسلة ثنائية، كل منها بطول 132 بت. تمثل كل سلسلة مفردة 22 وسيطاً كل منها بطول 6 بتات. يوافق كل من وسطاء المتحولات مكاناً على المحور x لأساس تابع الانتماء. مثلاً، تلزم 6×6 بتاً لتعريف توابع الانتماء الأربعة لارتفاع السائل h على النحو التالي: 6 خانات استعملت لتعريف مكان على المحور x لسطر تابع الانتماء "منخفض"، وتوافق 6 خانات "عال"، و12 خانة لمكاني أساس لـــ" منخفض متوسط"، و12 خانة من أجل "عال متوسط".

وبالمثل يتطلب كل من توابع الانتماء الخمسة في حالة *dh/dt* و$Q_{in}$ ($Q_{out}$) 8×6 خانة لتمثيل قيم الوسطاء.

قورنت نتائج المحاكاة مع التجارب التقليدية، فأثبتت حلول الخطأ أن حلول الخوارزميات الوراثية لتصاميم تابع الانتماء كانت أعلى في جميع الحالات. قادت عناصر التحكم المصممة بخوارزميات وراثية ارتفاع السائل إلى وضعه عند النقطة المطلوبة بسرعة أكبر، وكانت المحافظة على المستوى هناك أكثر استقراراً من أي تقريب آخر.

## 4.15 الخوارزميات الوراثية والشبكات العصبونية

## Genetic Algorithms and Neural Networks

تعتبر الخوارزميات الوراثية والشبكات العصبونية خوارزميات استمثالية. تبحث الخوارزميات الوراثية في إيجاد عضو، أو أعضاء كثيرة، من المجتمع الذي يمثل حلاً للتابع الموضوعي، أما خوارزميات تعليم الشبكات العصبونية الصنعية فإنها تبحث في إيجاد مجموعة من الأوزان التـــي تقلل عدد التصنيفات غير الصحيحة. الاثنان أيضاً مرتبطان بحاسة واحدة،

حيث كلاهما نموذج لعمليات عصبونية تتكيف لتحسين الإنجاز، ومن ثم زيادة فرصها في البقاء. صحيح أنهما يشتركان في هذه التجارب العامة، ولكن تركيبهما يستعمل في التطبيقات المرتبطة أكثر باستعمال الخوارزميات الوراثية في بناء بنـــى أفضل للشبكات العصبونية الصنعية أو في تهذيب وتحسين إنجاز بنـــى معطاة. من وجهة النظر هذه، استعملت الخوارزميات الوراثية في تصاميم الشبكات العصبونية للأغراض التالية:

1. إيجاد مجموعة استمثالية للأوزان في الشبكات العصبونية الصنعية المتعددة الطبقات الأمامية التغذية (بحجم أصغر) (Montana وDavis عام 1989[246]، وWhitley عام 1993[247]).
2. بناء شبكات استمثاليـــة لمسائـــل معطـــاة (Bornholdt وGraudenz عام 1992[180]، وMandischer عام 1993[248]).
3. إيجاد وسطاء استمثالية للشبكة (معدل التعليم، ومقدار حد كمية الحركة، وعدد العقد) لمسائل معطاة (Murray علم 1994[158]).
4. إيجاد مجموعة القواعد التـــي تصف سلوك الشبكة المدربة (Mitchll عام 1993 [249]).

في معظم الحالات المذكورة آنفاً، المشكلة الحقيقية هي في إيجاد تمثيل مناسب للمتحولات وتحويله إلى فراغ حل بشروط مقيدة، وهذا يجعل الأمر طيعاً لحل الخوارزميات الوراثية.

ومع أن بعض النتائج كانت مشجعة، فما يزال هناك عمل كثير يجب فعله قبل الاستفادة الكاملة من الخوارزميات الوراثية المصممة لدعم الجهود في تطوير الشبكات العصبونية الصنعية.

## 5.15 لمحة عن شبكات المنطق العصبونـــي

## Overview of Neural Logic Networks(NLNs)

شبكات المنطق العصبونـــي هي شبكات لمعالجات بسيطة، تستعمل تمثيلاً منطقياً ثلاثي القيمة: صح (true)، وخطأ (false)، وغير معروف unknown (نعم (yes)، ولا (no)، ولا أعلم (don't know)). تضم هذه الشبكات الملامح البارزة للمعالجة التفرعية مثل التكيف والمنطق التقليدي.

طور Teh Hoon Heng هذه الشبكات عام 1989[250]، وChan عام 1991[161] في

معهد علم الأنظمة (Institute of Systems Science) التابع للجامعة الوطنية بسنغافورا، حيث كانت البداية عام 1980.

ومنذ نشر مقترحهم الأولي، بدأت أفكارهم تتوسع ليشمل مختلف بنـــى الشبكات، وخوارزميات التعليم، ودمج المنطق العائم في أنظمة المنطق العائم العصبوني، والتعليل الاحتمالي المنطقي العصبوني.

في هذه الفقرة، سنعطي فقط وصفاً مختصراً للمفاهيم الأساسية ولبعض البنـــى الصندوقية المستعملة لتطوير شبكات أكثر تقدماً (محنكة).

في تركيب الحساب العصبونـــي مع منطق جبر بول (Bool)، نكسب قدرة في معالجة النموذج والتعليل المنطقي ضمن نفس إطار العمل. ففي استعمال ثلاث قيم عوضاً عن قيمتـــي المنطق الكلاسيكي، نربح قوة تعبيرية أكثر في تمثيل المعرفة والتبسيط في عملية التعليل. يمكن أن تلخص العمليات الأساسية لشبكات المنطق العصبونـــي كما يلي:

شبكات المنطق العصبونـــي هي مخططات مباشرة مؤلفة من دخل وخرج وعقد مخفية (رؤوس) مع وصلات مباشرة (أطراف). المداخل وقيم تفعيل العقد كلها ثلاثية القيمة بتمثيلات أزواج مرتبة كما يلي:

(1, 0) مقابل صح

(0, 1) مقابل خطأ

(0, 0) مقابل لا أعلم

الوصلات بين العقد لها أوزان بأزواج $(v,w)$ مرتبة بقيم حقيقية (موجبة، أو سالبة، أو أصفار)، والتـــي يمكن أن تعلم كيفياً، أو يمكن أن تحسب مباشرة لمسائل خاصة. تكون الأوزان الموجبة مهيجة وتكون الأوزان السالبة مخمدة. ينفذ الاستدلال بواسطة إرسال الإشارات باتجاه التغذية الأمامية ويعتمد على الحسابات التالية:

لتكن S عقدة بمداخل $(\alpha_i,\beta_i)$ من العقد $V_1,V_2,...,V_k$، بأوزان الوصلة الموافقة $(v_i,w_i)$، وعتبة تفعيل مساوية للواحد:

1. حساب:

$$net = \sum_{i=1}^{k} \left( \alpha_i v_i - \beta_i w_i \right)$$

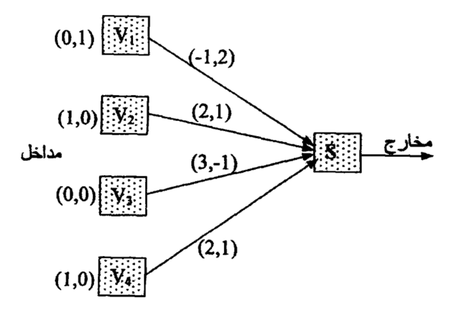
2. وضع تابع التفعيل للعقدة S وفقاً لـــ:

$$(\alpha_s, \beta_s) = \begin{cases} (1,0) & net \geq 1 \\ (0,1) & net \leq -1 \\ (0,0) & otherwise \end{cases}$$

كمثال عـــن المفاهيم السابقة، نأخذ شبكة المنطـــق العصبونـــي الموضحة فـــي الشكل (6.15). جرى في هذه الشبكة البسيطة (k = 4) حساب تابع التفعيل كما يلي:

$$\begin{aligned} (0, 1) \times (-1, 2) &= (0, 2) \\ (1, 0) \times (2, 1) &= (2, 0) \\ (0, 0) \times (3, -1) &= (0, 0) \\ (1, 0) \times (2, 1) &= (2, 0) \\ &= (4, 2) \end{aligned}$$

وهكذا net = 4–2، ومن ثم يكون تابع التفعيل لـــ S هو (1,0).




**الشكل 6.15:** شبكة منطق عصبونـــي بأربعة مداخل

باستعمال قواعد الانتشار السابقة، نستطيع بسهولة بناء شبكات المنطق العصبونـــي التـــي تنفذ توابع جبر بول، مثل OR، وAND، وNOT، والتضمين (Implication) الموضحة في الشكل (7.15)، وتابع XOR، والأكثرية (Majority) الموضحة في الشكل (8.15) ونترك للقارئ العزيز مقارنة هذه الشبكات مع مثيلاتها التـــي تنفذ نفس التواع المنطقية المبنية من العصبونات الصنعية البسيطة المعطاة في الفصول الأولى من هذا الكتاب.

لاحظ الازدواجية بين OR وAND. وهناك عمليات شبكات منطق عصبونـــي أخرى لها هذه الملامح المزدوجة مثل XOR وXAND ...الخ. العدد الكلي للعمليات الأساسية هو 28 عملية يمكن أن يركب بعضها مع بعض لتعطي مجالاً واسعاً من التعابير.

مثلاً، شبكات المنطق العصبونـــي البسيطة الموضحة في الشكل(7.15) يمكن أن يركب بعضها مع بعض لتصبح شبكات أكثر تعقيداً لتشكيل قاعدة في نظام خبير. لنأخذ قاعدة مع سابقة مؤلفة من فاصلين (ليسا رابطين) شرطيين، الأول تعبير مرفوض ورابطين آخرين والمحصلة المفردة. يمكن أن تكتب مثل هذه القاعدة كما يلي:

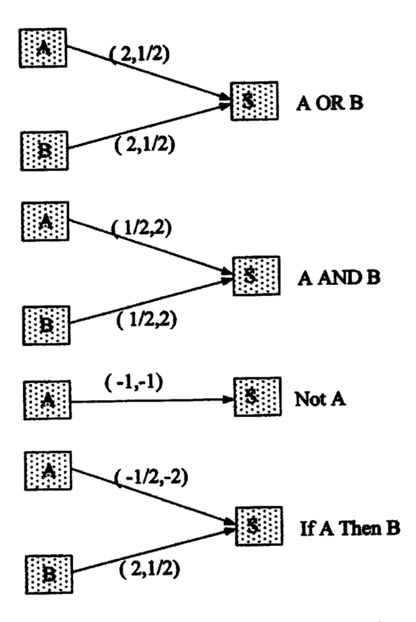
**إذا كان**

(تعبير-*1* وتعبير-*2* ) أو ( **ليس** تعبير-*3*)

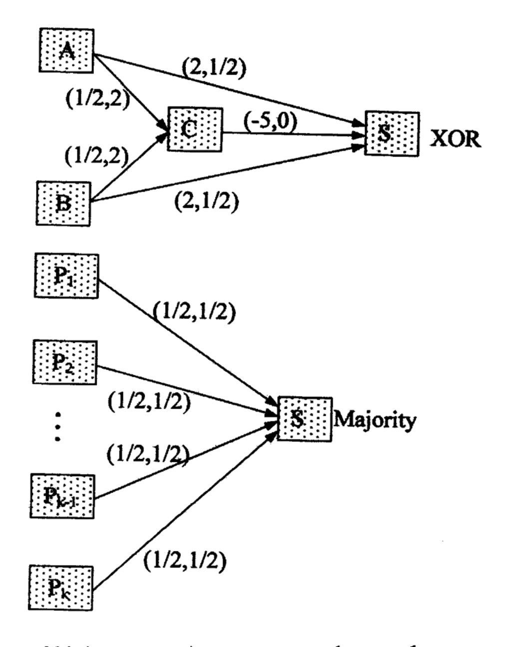
**فإن** (تعبير-4)

If (expression-1 AND expression-2) OR (NOT expression-3) Then (expression-4)

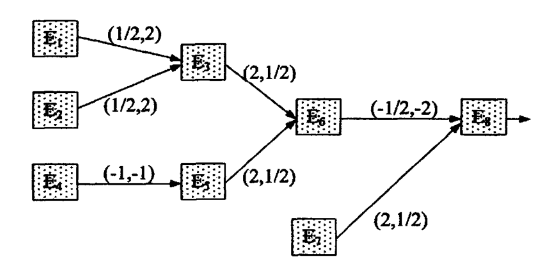
ويمكن أن تبنـــى مـــن بناء صندوقي بالاستعانة بالشكل (7.15) كما هو مبيـــن بالشكل (9.15). قارن هذه الشبكات مع مثيلاتها المعطاة في الفصول الأولى من الكتاب.




**الشكل 7.15**: شبكات المنطق العصبوني البسيطة لتنفيذ توابع جبر بول




الشكل8.15: شبكات المنطق العصبوني لحساب XOR وMajority




**الشكل 9.15**: قاعدة نظام خبير إذا كان ($E_1$ و $E_2$) أو (ليس $E_4$) فإن $E_7$

باستعمال هذه الشبكة وشبكات البناء الصندوقي الأساسية، نستطيع تركيب قواعد تعبير بعدة حقائق مختلفة، ويمكن أن توجد العلاقات بينها. ويمكن إيجاد عدد من القواعد المتضمن حقائق مرتبطة بمجال خبرة معطى لتشكيل أساس قاعدة معرفة كاملة في نظام خبير. مثلاً، يمكن تركيب نظام خبير تشخيصي للدارات الإلكترونية المعقدة من هذه القواعد حيث تتضمن الافتراضات أعراض الخطأ الشرطية، وتعطي محصلات القواعد الأخطاء الموافقة للأعراض.

إذا كانت المعرفة مرتبطة بأخطاء ممكنة عديدة غير معروفة، وكانت نماذج الأعراض والأخطاء متوفرة، فإن الشبكة يمكن أن تدرب كما في حالة الشبكات العصبونية الصنعية الأخرى المدروسة فيما سبق من فصول.

يمكن أن يسوى تعليل غير مؤكد في شبكات المنطق العصبونـــي بتوسيع عملها لتشمل قيماً احتمالية. التقريب الأبسط لتحديد قيم الاحتمال a و b للشروط true وfalse، أي، (a, b) هو زوج تفعيل احتمالي يوافق القيم true وfalse، حيث $0 \le a \le 1$ و $0 \le b \le 1$ و $0 \le a+b \le 1$، إذاً: الاحتمال غير المعروف هو $1-(a+b)$.

أخيراً، ركبت شبكات المنطق العصبونـــي مع المنطق العائم لبناء شبكات المنطق العصبونـــي العائم. فـــي هذه الحالة، استعـــمل الزوج المرتب (a, b) مع $0 \le a+b \le 1$ ليمثل "علامة أو دليلاً" للشرط، أو عكس الشرط، أو فقدان العلامة فيما يتصل بالشرط.

تـــرمز القيمة "a" إلى علامة الشرط، وتـــرمز "b" إلى علامة عكس الشرط، وترمز (1−(a + b لفقدان العلامة فيما يتصل بالشرط. مثلاً، إذا جمع رأي 100 خبير فيما يخص وضعاً ما، وكان 75 منهم موافقاً (true)، و15 غيـــر موافـــق، و10 قالوا لا نعلم. عندئذ (a, b) = ( 0.75, 0.15).

لتدريب شبكات المنطق العصبونـــي يمكن أن تستعمل أي طريقة من طرائق تدريب الشبكات العصبونية الصنعية التقليدية، بما في ذلك الانتشار الخلفي وperceptron وقاعدة Delta البسيطة.

إن المعالجة الكاملة لنظرية شبكات المنطق العصبونـــي تقع خارج موضوع هذا الكتاب، وهي تتطلب كتباً عديدة لوصف هذا الموضوع جيداً.

في إدراك لهذا العمل وأهميته منحت الحكومة اليابانية معهد علوم الأنظمة جائزة البحث العلمي كجزء من مشروع حساب العالم الحقيقي المدعوم من MITI اليابانية.

## 6.15 توجهات مستقبلية في مجال الشبكات العصبونية

من الواضح أن تطورات جديدة آتية قريباً في إنجاز الشبكات العصبونية الصنعية من جهة التصميم الصلب (Hardware). عبر الثلاثين سنة الماضية، تقلص حجم التجهيزات (العتاد) الصلبة إلى النصف تقريباً كل ثلاث سنوات، وبنفس الوقت تضاعفت السرعة تقريباً خلال هذه المدة، وانخفضت الكلفة المالية انخفاضاً مثيراً في هذه المدة.

حالياً يمكـــن وضـــع عشـــرات الملايـــن مـــن الترانـــزستورات على شريحة VLSI؛ (Very Large Scale Integrated) واحدة. من المحتمل أن تطورات إنقاص الحجم مع زيادة سرعات الحساب ستستمر بهذا المعدل حتـــى الوصول إلى الحدود الفيزيائية. عندها سيكون من الممكن بناء شبكات عصبونية صنعية بملايين العصبونات السيليكونية وبذلك يبدأ تقليد حقيقي لوظائف المعالجة البشرية الذكية، ومع الزيادة في الحجم سيكون هناك زيادة في التعقيد. عندها سنرى بنـــى جديدة للشبكات العصبونية الصنعية مبنية من شبكات هجينية تدمج وحدات تعرف الأشكال بمستوى منخفض مع وحدات معالجة المستوى العالي لتقوم بإنجاز مهام متعددة الوظائف مشابهة لمعالجة الرؤية البشرية، لكن بأسلوب أبسط.

على أية حال، من المحتمل أن يأتـــي التقدم الحقيقي من التصنيع لأدوات الجزيء الحيوي (biomolecular) التجارية، التـــي بدأت بالظهور في الأسواق في بداية القرن الحادي والعشرين. مثل هذه الأدوات يلزم لمحاكاة سعات التخزين الهائلة ومقدرات الحساب المتوازية الضخمة لأجزاء من الدماغ البشري. هذا هو التحدي الهائل أمام الباحثين في مجال الشبكات العصبونية الصنعية. يجري العمل على قدم وساق في عدد من مخابر الولايات المتحدة واليابان وأوروبا (ونتساءل أين مخابر العالم العربي) لبناء معالجات الجزيء الحيوي وأدوات التخزين، وقد تحققت بعض النجاحات المذهلة.

ليس جزافاً الاعتقاد أننا سنرى بعض الشبكات العصبونية **الطبيعية الصنعية** في مطلع القرن الحادي والعشرين. طبعاً هذه الشبكات ستكون ميكروسكوبية مقارنة مع الدماغ البشري، ووظائفها محدودة. ومع ذلك، يمكن أن تظهر شبكات قوية وستصبح حقيقة بمقياس عالٍ يصل إلى أحجام الدماغ المصغر (minibrain). وحيث إن النجاحات الأولية في تطورات الجزيئات الحيوية قد حصلت بالفعل، بل ذُكرت في الكتاب الحديث لـــ D. H. Freedman الذي يحمل اسم Brainmakers المنشور عام 1994[196]، والتـــي أحجمنا عن ذكرها هنا، فإننا نضعها نصب عينـــي القارئ المهتم لكي يقتنع بالاتجاه الحتمي لأبحاث الشبكات العصبونية الصنعية التـــي ستقودنا في المستقبل القريب شئنا أم أبينا.

# دليل المصطلحات العلمية

## إنكليزي-عربي

-A-

| | |
|---|---|
| Activation | تفعيل |
| Adaptive | متكيف |
| Algorithm | خوارزمية |
| Ambiguous | غامضة |
| Analog-digital | تماثلي رقمي |
| Antiderivative | عكس المشتق |
| Argument | محاكمة ــ مُحدِّد |
| Artificial | صنعي |
| Association | اقتران- ترافق |
| Associative recall | استدعاء مترافق |
| Associative property | خاصة تجميعية (اقترانية) |
| Attractor | جاذب |
| Autoassociative | ترافق ذاتي |
| Autocorrelation | ارتباط ذاتي |
| Autonomous | ذاتي القيادة |
| Axon | محور |

-B-

| | |
|---|---|
| Backpropagation | انتشار خلفي (عكسي) |
| Basin of attraction | حوض التجاذب |
| Batch learning | تعليم دُفعّي |
| Bias | انحياز |
| Bidirectional | ثنائي الاتجاه |
| Binomial | ثنائي الحد |
| Biology | بيولوجيا ــ علم الحياة |
| Biomolecular | جزئ حيوي |
| Bipolar | ثنائي القطبية |

| | |
|---|---|
| Bit | بت ـــ خانة ثنائية |
| Brain | دماغ |

-C-

| | |
|---|---|
| Cardinality | رئيسي / أصلي |
| Cascade | متتابع |
| Category | فئة |
| Cell | خلية |
| Chain rule | قاعدة السلسلة |
| Chaos | فوضوي |
| Characteristic | مميِّز |
| Chip | شريحة ـــ رقائق |
| Chunking | تكتل |
| Clamped mode | نمط الإلزام |
| Class | صف |
| Classification | تصنيف |
| Cluster | قطاع ـــ عُنْقود (عناقيد) |
| Code | رمز |
| Combinatorial | تركيبي/ توافيقي |
| Compact | متراص |
| Comparator | مقارن |
| Competitive | تنافسي |
| Complement | مُتَمِّم |
| Commutative | تبديلي |
| Computer vision | رؤية حاسوبية |
| Concentration | تركيز |
| Conditional | شَرطيّ/ مشروط |
| Conditional connective | رابطة شرطية |
| Conservative | محافظ |
| Constraint | قَيْد ـــ شرط مقيّد |
| Content-addressable memory | ذاكرة معنونة بالمحتَوى |
| Context layer | طبقة السياق |

| | |
|---|---|
| Continuous | مستمر |
| Contradiction | تناقض |
| Control | تحكم/ضبْط |
| Controller | مُتحكم |
| Convergence | تقارب |
| Converter | مُحوّل/ مبدل |
| Correlation | ترابط |
| Coulomb | كولومب |
| Counterpropagation | الانتشار المتعاكس |
| Column | عمود |
| Covariance | تباين مشترك |
| Credit-assignment | تخصيص الاعتماد |
| Crossover | تقاطع/ تصالب |
| Cumulative | تراكمي |

-D-

| | |
|---|---|
| Data compression | ضغط المعطيات |
| Decision region | منطقة القرار |
| Decoder | مُفكّك ترميز |
| Definite integral | تكامل محدود |
| Dendrite | فرع شجري |
| Derivative | مشتق |
| Detection | كشف |
| Diagonal matrix | مصفوفة قطرية |
| Diagnosis | تشخيص |
| Diagram | مخطط |
| Differential equation | معادلة تفاضلية |
| Dilation | تمدد |
| Dimension | بُعْد |
| Discrete | مُقطّع/ منفصل |
| Dissipative | مبدد |
| Distribution | توزيع |

| | |
|---|---|
| Distributive | توزيعي |
| Disturbance | اضطراب |

-E-

| | |
|---|---|
| Echo cancellation | حذف الصدى |
| Eigenvalue | قيمة خاصة |
| Eigenvector | شعاع خاص |
| Electrocardiogram | مخطط، تخطيط القلب |
| Electrochemical | إلكتروكيميائية |
| Electrode | إلكترود |
| Electronic switching | تبديل إلكتروني |
| Element | عنصر |
| Embedding | تضمين |
| Empty set | مجموعة خالية |
| Emulation | محاكاة ـــ مضاهاة/ تقليد |
| Encoding | ترميز |
| Encoder | مُرَمِّز |
| Energy | طاقة |
| Entropy | أنتروبي |
| Epoch | دور |
| Equalization | موازنة ـــ تسوية |
| Equation | معادلة |
| Estimation | تقدير |
| Estimator | مُقَدِّر |
| Erasing | المحو ـــ الحذف |
| Euclidean | الإقليدي |
| Excitatory | مُهيِّج/ مُحَرِّض |
| Expectation value | قيمة متوقعة |
| Expert system | نظام خبير |
| Exponent | أُسّ |
| Expression | تعبير |
| Extrapolation | استيفاء خارجي |

-F-

| | |
|---|---|
| Far-end | النهاية البعيدة |
| Fault | خطأ/ عطل/ خلل |
| Feature | سمة/ ميزَة |
| Feature maps | خرائط سمات |
| Feedback | تغذية راجعة |
| Feedforward | تغذية أمامية |
| Filtering | ترشيح |
| Fitting | إلباس |
| Flow | تدفق انسياب/ جريان |
| Forecasting | تنبؤ |
| Format | مَصَاغَة |
| Function | وظيفة/ إجراء |
| Fusion | دمج |
| Fuzzy | ترجيجي ـــ عائم ـــ مبهم |

-G-

| | |
|---|---|
| Generalization | تعميم |
| Genetic algorithm | خوارزمية جينية |
| Global minimum | أصغر شامل/ شمولي |
| Gradient descent | تدرج الهبوط |
| Grid | شَبَكْ |
| Grouping | تجمّع |

-H-

| | |
|---|---|
| Heteroassociative | ترافق مغاير |
| Heuristic method | طريقة تجريبية كسبية |
| Hidden layer | طبقة مخفية |
| Histogram | مخطط [بياني] نسيجي |
| Hyberbolic | زائدي المقطع |
| Hyberbolic tangent | ظل قطعي |
| Hybrid | هجين |

-I-

| | |
|---|---|
| Idempotent | اللانمو |
| Identification | تعريف |
| Identifier | مُعرِّف/ مُعيِّن [الهوية] |
| Identity matrix | مصفوفة واحدية |
| Identity function | تابع التماثل |
| Image processing | معالجة الصور |
| Implication | التضمين |
| Indefinite integral | تكامل غير محدود |
| Index | فهرس/ دليل |
| Inequality | متراجحة |
| Inference | استدلال |
| Information theory | نظرية المعلومات |
| Inhibitory | مُخمِّد |
| Initial value | قيمة أولية |
| Input | دخل |
| Integral | تكامل |
| Intensity | شدة |
| Interpolation | استيفاء داخلي– توليد |
| Interval | مجال |
| Invariant | لاتغيري |
| Inversion | قلب/ عكس |

-J-

| | |
|---|---|
| Joint | مشترك |

-K-

| | |
|---|---|
| Kinetic energy | طاقة حركية |

-L-

| | |
|---|---|
| Layer | طبقة |
| Learning | تعليم |
| Least squares | مربعات صغرى |
| Limit | نهاية/ حد |

| | |
|---|---|
| Linear | خطي |
| Linear independent | مستقل خطياً |
| Local minimum | أصغر محلي |
| Long term | أجل طويل |

-M-

| | |
|---|---|
| Manifold | جملة مولدة |
| Mapper | مطبق |
| Mapping | تطبيق /مقابلة/ إسقاط [طباقياً] |
| Marginal | هامشي |
| Mating | تزاوج |
| Matrix | مصفوفة |
| Maximization | تعظيم/ تكبير |
| Mean | متوسط/ وسطي |
| Mean square error | متوسط مربع الخطأ |
| Median | وسط، الفاصل في الوسط |
| Membrane potential | كمون الغشاء |
| Memorization | تذكر (وضع في الذاكرة) |
| Metric | مسافة |
| Mode | نمط |
| Model | نموذج |
| Modeling | نمذجة |
| MODEM | مودم: تعديل/فك تعديل |
| Moment | عزم |
| Momentum | كمية حركة |
| Monitoring | مراقبة |
| Multilayer | متعدد الطبقات |
| Multisensor | متعدد الحساسات |
| Mutation | طفرة وراثية |

-N-

| | |
|---|---|
| Near-end | النهاية القريبة |
| Neuro-fuzzy | عصبوني عائم |

| | |
|---|---|
| Neuro-logic | منطق عصبوني |
| Neuron | عصبون |
| Network | شبكة |
| Node | عقدة |
| Noise | ضجيج |
| Nonlinear | لاخطي |
| Nonparametric | موسطات |
| Norm | نظيم |
| Normal distribution | توزيع نظامي |
| Normalization | استنظام |
| Nucleus | نواة |

-O-

| | |
|---|---|
| Object | غرض |
| Off-line learning | تعليم مفصول (مؤجل) غير مباشر |
| Offspring | نسل |
| On-line | موصول [إلى الخط] مُتاح |
| Optimization | استمثال |
| Orbit | مدار |
| Orthogonal | متعامد |
| Outer product | جداء خارجي |
| Output | خرج |

-P-

| | |
|---|---|
| Parallel processing | معالجة متوازية |
| Parameter | مُوسط |
| Parity | ندّيَة، ندية |
| Pattern | نموذَج- شكل |
| Pattern recognition | تعرّف الأشكال [النمطية] |
| Percentile | نسبة مئوية |
| Perception | إدراك |
| Performance | أداء - إنجاز |
| Phase | طور |

| | |
|---|---|
| Photoreceptor | مستقبل ضوئي |
| Pixel | بكسل (عنصر صورة) |
| Plasticity | اللدونة |
| Principal component | مركبة أساسية |
| Pocket | محفظة |
| Polynomial | كثير حدود |
| Prediction | تنبؤ |
| Probability | احتمال |
| Process | اجرائية |
| Processing | معالجة |
| Prototypical | نمذجة أولية |
| Pseudo-inverse | معكوس مُفترض |

-Q-

| | |
|---|---|
| Quantization | استكمام |
| Quantizer | مكمم |
| Quickprop | الانتشار السريع |

-R-

| | |
|---|---|
| Radial | شعاعي |
| Random | عشوائي |
| Random variable | متحول عشوائي |
| Range | مجال |
| Rank | رتبة |
| Rate | معدل |
| Recurrent | تكراري |
| Regression | انكفاء |
| Reinforced | معززة |
| Reinforcement | تقوية-تعزيز |
| Relative | نسبي |
| Relaxation | استرخاء |
| Representation | تمثيل |
| Resonance | طنين |

| | |
|---|---|
| Response | استجابة |
| Retrieval | استحضار |
| Risk | مجازفة |
| Robust | متين/ منيع (حصين) |
| Row | سطر |
| Rule | قاعدة (ناظمة) |

-S-

| | |
|---|---|
| Scalar product | جداء سُلّمي |
| Schedule | يُجدوِل (زمنياً) [جدول زمني] |
| Segment | مُقْتَطَع |
| Selection | تحديد/ انتقاء |
| Self-adaptive | متكيف ذاتياً |
| Self-growing | نمو ذاتي |
| Sensor | مُحس |
| Servo | [محَرك] تخديم |
| Set | مجموعة |
| Shift register | سجل إزاحة |
| Short term | أجل قصير |
| Signal | إشارة |
| Simulated annealing | محاكاة التلدين |
| Simulated networks | شبكات المحاكاة |
| Simulation | محاكاة |
| Soft computing | حساب لين |
| Soma | جسم الخلية |
| Spectrum | طيف |
| Stability | الاستقرار |
| Stable | مُستقر |
| Standard deviation | انحراف معياري/انحراف قياسي |
| State | حالة |
| Statistical | إحصائي |
| Stimulus | منبه |

| | |
|---|---|
| Stochastic | عشوائي |
| Storage | وسيطه خزن (خزّان) |
| Structure | بنية |
| Supervised –learning | تعليم بمعلم |
| Synapse | ليف |
| Synchronous | متزامن |
| System | نظام |

-T-

| | |
|---|---|
| Target output | الخرج الهدف (أو المنشود) |
| Term | حد-أجل |
| Theory | نظرية |
| Threshold | عتبة |
| Tolerance | تسامح |
| Toutology | كامل |
| Training | تدريب |
| Trajectory | مسار |
| Transformation | تحويل |
| Transformer | مُحوّل |
| Transitive | متعدية |
| Transpose | منقول |
| Truth table | جدول الحقيقة |

-U-

| | |
|---|---|
| Unambiguous | غير غامضة |
| Uncertainty | شك |
| Unclamped mode | نمط عدم الإلزام |
| Unit | وحدة |
| Universal set | مجموعة عميمة |
| Unstable | غير مستقر |

-V-

| | |
|---|---|
| Valid argument | محاكمة صحيحة |
| Value | قيمة |

| | |
|---|---|
| Variance | تباين |
| Vector | متجه |
| Vigilance | يقظة-احتراس |
| Voxel | عنصر حجم |

-W-

| | |
|---|---|
| Weight | وزن/ ثقَل |
| Weighted sum | مجموع مُثقّل |

# المراجـــع


[1]- Hebb, D. O. *"The Organization of Behavior"* Wiley, New York, 1949.

[2]- Hopfield, J. J. & Tank, D. W.*"Neural Computation of Decisions in Optimization Problems"* Biological Cybernetics,Vol.52, pp.141-152, 1985.

[3]- Patterson, D. W., Chan, K. H., Tan, C. M.*"Time Series Forecasting with Neural Networks: A Comparative Study"* Proceedings of the International Conference on Neural Networks, Applications of signal processing (NNASP-93), Singapore, pp.269-274, 1993.

[4]- Zak, M. *"An Unpredictable - Dynamics Approach to Neural Intelligence"* IEEE Expert, pp.4-10, August, 1991.

[5]- Lippman, R. P.*"An Introduction to Computing with Neural Nets"* IEEE ASSP Magazine, pp.4-22, April, 1987.

[6]- Linsker, R. *"Self-Organization in a Perceptual Network"* IEEE Computer, Vol.21, No.3, pp. 105-117, 1988.

[7]- Kohonen, T.*"Self-Organized Formation of Topologically Correct FeatureMaps"* Biological Cybernetics, Vol.43, pp.59- 69, 1982.

[8]- Fukushima, K. & Miyake, S.*"Necognitron:A New Algorithm for Pattern Recognition Tolerant of Deformation and Shifts in Position"* Pattern Recognition, Vol.15, No.6, pp. 455 - 469, 1982.

[9]- Grossberg, S. *"Nonlinear Neural Networks: Principles, Mechanisms, and Architectures"* Neural Networks, Vol.1, No.1, pp. 17-61, 1988.

[10]- Hopfield, J. J. *"Neural Networks and Physical Systems with EmergentCollective Computational Abilities"* Proceedings of the National Academy of sience,Vol.79, pp. 2554-2558, 1982.

[11]- Anderson , J. A. *"Cognitive and Psychological Computation with Neural Models"* IEEE Transactions on Systems, Man, and Cybernetics,Vol. SMC-13, No.5, pp.799 - 815, 1983.

[12]- Kosko, Bart *"Adaptive Bidirectional Associative Memories"* Applied Optics, Vol.26, No.23, pp.4947-4960, 1987.

[13]- Shannon, C. E. *"A Mathematical Theory of Communication"* Bell System Technical Journal, Vol.27, pp.379-423, pp.623 - 656, 1948.

[14]- Gutzwiller, M. C. *"Quantum Chaos"* Scientific American, pp.78-84, January, 1992.

[15]- Lorentz, E. N. *"Computational Chaos-A Prelude to Computational Instability"* Physica D, Vol.35, pp.299-317, 1989.

[16]- Takens, F. *"Detecting Strange Attractors in Turbulence, in Dynamical Systems and Turbulence"* Warwick, 1980, Lecture Notes in mathematics No.898,Rand, D. and Young, L. S., eds., Springer, Berlin, pp.366-381,1981.

[17]- Mane, R. *"Dynamical Systems and Turbulence"* Warwick, 1980, Lecture Notes in mathematics No.898, Rand, D. and Young, L. S., eds., Springer, Berlin, pp.230-242, 1981.

[18]- Ababarbanel, H. D. I. & Reggie Browen, & Tsimring, L. S. *"The Analysis of Observed Chaotic Data in Physical Systems"* unpublished document preprint,1993.

[19]- McCulloch, W. S. & Pitts W. *"A Logical Calculus of the Ideas Immanent in Nervous Activity"* Bulletin of mathematical Biophysics, Vol.5, pp.115-133, 1943.

[20]- Rumelhart, D. & McClelland, J. *"Parallel Distributed Processing"* Vol.1, eds., MIT press, Cambridge, MA, 1986.

[21]- Rosenblatt, F. *"The Perceptron: Probabilistic Model for Information Storage and Organization in the brain"* Psychological Review, Vol. 65, pp.386-408. Reprinted in Anderson & Rosenfeld [1988], pp.92-114, 1958.

[22]- Gallant, S. I. *"Neural Network Learning and Expert Systems"* MIT Press, Cambridge, MA, 1993.

[23]- Hertz, J. A. & Krogh, A. & Palmer, R. G. *"Introduction to the Theory of Neural Computation"* Addison-Wesley, Redwood City, CA, 1991.

[24]- Minsky, M. L. & Papert, S. A. *"Perceptrons"* Expanded edition, Cambridge, MIT Press, MA, 1988.

[25]- Arbib,M. A. *"Brains, Machines, and Mathematics"* 2th ed., Springer-Verlag, New York, 1987.

[26]- Rosenblatt, F. *"Principles of Neurodynamics"* New York, Spartan, 1962.

[27]- Burr, D. J. "*An Improved Elastic Net Method for the Traveling Salesman Problem*" Proceedings of the IEEE International Conference on Neural Networks, San Diago, Vol.1, pp.69-76, 1988.

[28]- Patterson, D. W. "*Artificial Neural Networks: Theory and Applications*" Prentice Hall, Simon & Schuster(Asia), Singapore,1996.

[29]- Widrow, B. & Hoff, M. E. "*Adaptive Switching Circuits*" IRE WESCON Convention Record, New York, 1960.

[30]- Kohonen, T. "*Associative Memory :A System Theoretic Approach*" Springer, New York, 1977.

[31]- Stone, G. "*Parallel Distributed Processing*" Vol.1, MIT Press, MA, Cambridge, 1986.

[32]- Widrow, B. & Winter, R. "*Neural Nets for Adaptive Filtering and Adaptive Pattern Recognition*" IEEE Computer, Vol.21, No.3, pp.25-39, 1988.

[33]- Andes, D. & Widrow, B. & Leher, M. & Wan, E. "*MRIII: A Robust Algorithm for Training Analog Neural Networks*" Proceedings of the International Joint Conference on Neural Networks, Seattle, WA, Vol.I, Erlbaum, Hillsdale, NJ, pp.533-536, 1991.

[34]- Widrow, B. & Winter, R. G. & Baxter, R. A. "*Learning Phenomena in Layered Neural Networks*" Proceedings of the First IEEE International Conference on Neural Network, San Diago, 1987.

[35]- Widrow, B. & Rumelhart, D. & Lehr, M. "*Neural Networks : Applications in Industry, Business and Science*" Communication of the ACM, Vol.37, No.3, pp.93-105, 1994.

[36]- Tesauro, G. "*Simple Neural Models of Classical Conditioning*" Biological Cybernetics, Vol.55, pp.187-200, 1986.

[37]- Szu, H. H. "*Neural Networks: Theory, Applications and Computing*" Lecture Notes for UCLA Engineering Short Course, Engineering 819.185, March 20-23, 1989.

[38]- Anderson, J. A. "*A Simple Neural Network Generating an Interactive Memory*" Mathematical Biosciences, Vol.14, pp.197-220. Reprinted in Anderson & Rosenfeld [1988], pp.181-192, 1972.

[39]- Hopfield, J. J. *"Neurons with Graded response Have Collective Computational Properties Like Those of two-State Neurons"* Proceedings of the National Academy of science, Vol.81, pp.3088-3092, 1984.

[40]- Cohen, M. Y. & Grossberg, S. *"Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Competitive Neural Networks"* IEEE Transectios on Systems, Man and Cybernetics, Vol.13, pp.815-826, 1983.

[41]- Bilbro, G. & Miller, T. K. & Snyder, W. E.& Van den Bout, D. E. & White, M. *"Optimization by Men Field Annealing, in Advances in Neural Information Processing Systems"* T.D.S. Touretaky, Morgan Kaufmann, San Mateo, CA, pp.91-98, 1988.

[42]- Kosko, Bart *"Bidirectional Associative Memories"* IEEE transactions on systems, Man and Cybernetics, Vol.SMC-18,pp.49-60, 1988-a.

[43]- Kosko, Bart *"Feedback Stability and Unsupervised Learning"* Proceedings of the IEEE International Conference on Neural Networks, San diago,Vol.1, pp.141-149, 1988-b

[44]- Abu-Mostafa, Y. S. & St Jacques, J. –M. *"Information Capacity of the Hopfield Model"* IEEE transactions on Information Rheory, IT-31, pp.461-464, 1985.

[45]- McEliece, R. J. & Posner, E. C & Rodemich, E. R. & Venkatesh, S. S. *"The Capacity of the Hopfield Associative Memory"* IEEE Transactions on Information Theory, IT-33, pp.461-482, 1987.

[46]- Hecht-Nielsen, R. *"Neurocomputing"* Readings, MA, Addison-Wesley, 1990.

[47]- Anderson, J. A. & Silverstein, J. W. & Ritz, S. A. & Jones, R. S. *"Distictive Features, Categorical Perception, and Probability Learning: Some Applications of a Neural Model"* Psychological Review, Vol.84, pp.413-451, 1977.

[48]- Hecht-Nielsen, R. *"Applications of Counterpropagation Networks"* Neural networks,Vol. 1(2), pp.131-139, 1988.

[49]- Kosko, B. *"Bidirectional Associative Memories"* IEEE Transactions on Systems, Man, and Cybernetics, Vol.18, pp.49-60. Reprinted in Anderson, Pellionize & Rosenfeld[1990], 1988.

[50]- Rosenblatt, F. *"Principles of neurodynamics: Perceptrons and the Theory of Brain Mechanisms"* Spartan Books, Washington, DC, 1961.

[51]- Werbos, P. J. *"Beyond Regression: New Tools for prediction and Analysis in the Behavioral sciences"* PH.D. Thesis, Harvard University, 1974.

[52]- Parker, D. B. *"learning Logic"* Technical Report TR-47, Center for Computational Research in Economics and Management Science, Massachusetts Institute of Technology, Cambridge, MA, 1985.

[53]- Rumelhart, D. J. & Zipser, D. *"Feature Discovery by Competitive Learning"* Cognitive Science, Vol.9, pp.75-112, 1985.

[54]- Widrow, B. & Stearns, S. D. *"Adaptive Signal Processing"* Prentice-Hall, Englewood Cliffs, NJ, 1985.

[55]- Cottrell, G. W. & Munro, P. & Zipser, D. *"Image Compression by Back Propagation: An Example of Extensional Programming"* In N. E. Sharkey, ed., Models of Cognition: A Review of Cognitive Science, Norwood, NJ: Ablex Publishing Corp, pp.208-240, 1989.

[56]- Rumelhart, D. E. & Hinton, G. E. 7 Williams, R. J. *"Learning Internal Representation by Error propagation, in Parallel Distributed Processing: Explorations in the Microstructures of Cognition"* Vol.1, Foundations, Rumelhart, D. E. & McClelland, MIT Press, Cambridge, MA, 1986.

[57]- Arozullah, M. & Namphol, A. *"A Data Compression System Using Neural Networks; Based Architecture"* International Joint Conference on Neural Networks, San Diago , CA, I: 531-536, 1990.

[58]- Sonehara, N. & Kawato, M. & Nakane, K. *"Image Data Compression Using a Neural Network Model"* International Joint Conference on Neural Networks, Washington, DC,Vol. II, pp.35-41, 1989.

[59]- Yu, X. H. *"On the Nonexistence of Local Minima of the Backpropagation Error Surfaces"* Proceedings of the IJCNN-91, Singapore, Vol.2, pp.1272-1277, 1991.

[60]- Jacobs, R. A. *"Increased Rates of Convergence Through Learning Rate Adaptation"* Neural Networks, Vol.1, No.4, pp.295-307, 1988.

[61]- Fahlman, S. E. *"An Empirical Study of Learning Speed in Backpropagation Networks"* Carnegie Mellon Report No.CMU-CS-88-162, 1988.

[62]- Hirose, Yoshio, Koichi Yamashita & Shimpei Hajiya *"Backpropagation Algorithm Which Varies the Number of Hidden Units"* Neural Networks, Vol.4, No. 1, pp.61-66, 1991.

[63]- Chen, C. L. 7 Nutter, R. S. *"Improving the training Speed of Three-Layer Feedforward Nets by Optimal Estimation of the Initial Weights"* Proceedings of the IJCNN-91, Singapore, Vol.3, pp.2063-2038, 1991.

[64]- Denueux, T. & Lengelle, R. & Canu, S. *"Initialization of Weights in a Feedforward Neural network Using Prototypes"* Proceedings of te ICANN-91, Espoo, Finland, pp.623-628, 1991.

[65]- Kim, Y. K. & Ra, J. B. *"Weight Value Initialization for Improving training Speed in the Backpropagation Network"* proceedings of the IJCNN-91, Singapore,Vol.3, pp.2396-2401, 1991.

[66]- Haario, H. & Jokinen, P. *"Increasing the Learning Speed of Backpropagation Algorithm by Linearization, in Artificial Neural Networks"* Proceedings of the ICANN-91, Espoo, Finland, Vol.1, Kohonen, T. & Makisara, K. & Simula, O. & Kangas, J., North-Holland, Amesterdam, pp.629-634, 1991.

[67]- Tollenaere, Tom *"SuperSAB: Fast Adaptive Backpropagation with Good Scaling Properties"* Neural Networks, Vol.3, pp.561-573, 1990.

[68]- Minai, A. A. & Williams, R. D. *"Backpropagation Heuristics: A study of the Extended Delta-Bar-Delta Algorithm"*Proceedings of the IJCNN-90, San diago, Vol.1, pp.595-600, 1990.

[69]- Sato, A. *"An Analytical Study of the Momentum Term in a Backpropagation Algorithm"* Proceedings of the ICANN-91, Espoo, Finland, pp.617-622, 1991.

[70]- Fogelman-Soulie, F; *"Neural Network Architectures and Algorithms: A perspective, Artificial Neural Networks"* Vol.1, Kohonen, T. & Simula, O. & Kangas, J., North-Holland, Amsterdam, pp.605-615, 1991.

[71]- Matsuoka, K. & Yi, J. *"Backpropagation Based on the Logarithmic Error Function and Estimation of Local Mimima"* Proceedings of the IJCNN-91, Singapore, Vol.2, pp.1117-1122, 1991.

[72]- Van ooyen, A. & Nienhuis, B. "Improving the Convergence of the Backpropagation Algorithm" Neural Networks, Vol.5, pp.465-471, 1992.

[73]- Battiti, R. *"First-and Second-Order Methods for Learning : Between Steepest Descent and Newton's Method"* Neural Cpmputation,Vol.4, No.2, pp. 141-166,1992.

[74]- Bishop, C. "*Exact Calculation of the Hessian Matrix for the Multilayer Perceptron*" Neural Computation,Vol.4,pp.491-501,1992.

[75]- Ballard, D. H. & Brown, C. M. "*Computer Vision*" Prentive-Hall, Englewood Cliffs, NJ, 1982.

[76]- Thorpe, C. & Hebert, M. & Kanade, T. & Shafer, S. "*Toward Autonomous Driving: The CMU NavLab*" IEEE Expert, pp.31-42, August, 1991.

[77]- Kanade, T. & Reed, M. L. & Weiss, L. E. "*New Technologies and Applications in Robotics*" Communications of the ACM, Vol.37, No.3, pp.58-67, 1994.

[78]- Staib, W. E. "*The Intelligent Arc Furance: Neural Networks Revolutionize Steel-Making*" Proceedings of INNS meeting, World Congress on Neural Networks, Portland, Or, Vol.I, pp.466 - 469, 1993.

[79]- Asakawa, Kazuo & Hideyuki Takagi "*Neural Networks in Japan*" Communications of the ACM,Vol.37, No.3, pp.106 - 12, 1994.

[80]- Rock, D. & Malkoff D. & Stewart, R. "*AI and Aircraft Health Monitoring*" AI Expert, pp.28-35, February, 1993.

[81]- Morose, R. A. "*A Financial Neural Network Application*" AI Expert, pp.50-53,May, 1990.

[82]- Morose, R. A."*A Financial Neural Network Application, in Neural Networks in Finance and Investing*" Trippi, R. R. & Turban, E., Probus Publishing, Chicago, pp.75-83, 1993.

[83]- le Cun. Y."*Models Connexionnistes de l'apprentissage*" Doctoral Dissertation, University of Pierre and Marie Curie, Paris, 1987.

[84]- le Cun, Y. & Boser, B. & Denker, J. S. & Solla, S. & Howard, R. & Jackel, L. "*Backpropagation Applied to Handwritten Zipcod Recognition*" Neural Computation, Vol.1, pp.541-551, 1990.

[85]- Keeler, J. D. & Rumelhart, D. E. & Leow, W. K. "*Integrated Segmentation and Recognition of Hand-Printed Numerals, in Neural Information Processing Systems*" Vol.3, Lippman, R. P. & Moody, J. E. & Touretzky, D. S., Morgan Kaufmann, San Mateo, CA, pp.557-563, 1992.

[86]- Lee, C. M. & Patterson, D. W. "*Occluded Object Recognition: An Approach Which Combines Neurocomputing and Conventional Algorithm*" Proceedings of the IJCNN-91, Singapore, Vol.2, pp.2612-2617, 1991.

[87]- Yanicoglu, B. A. & Sandon, P. A. *"off-Line Cursive Handwriting recognition Using Neural Networks"* Proceedings of the SPIE Applications of Artificial Neural Network IV, Orlando, FL, pp.102-106, 1993.

[88]- Hamilton & Hufnagel *"Early Detection of Epileptic Attacks, in Applications of neural Networks"* Schuster, H. G., VCH Verlagsgesellschaft, Weinheim, pp.173-178, 1992.

[89]- Colombi, J. M. & Anderson, T. R. & Rogers, S. K. *"Auditory Model Representation for Speaker Recognition"* Proceedings of the SPIE Applications of Artificial neural Networks IV,Orlando,FL,pp.9-14,1993.

[90]- Minsky, M. L. & Papert, S. A. *"Perceptrons"* MA:MIT Press,Original edition, 1969.

[91]- Pinda, F. J. *"Generalization of Backpropagation to Recurrent Neural Networks"* Physical Review Letters,Vol.59, pp.2229-2232, 1987.

[92]- Pinda, F. J. *"Dynamics and Architecture for Neural Computation"* Journal of Complexity,Vol.4, pp.216-245, 1988.

[93]- Pinda, F. J. "Recurrent Backpropagation and the Dynamical Approach to Adaptive Neural Computation" Neural Computation, Vol.1, pp.161-172, 1989.

[94]- Almeida, L. B. *"A Learning Rule for Asynchronous Perceptrons with Feedback in a Networks"* San Diago, Vol.2, pp.609-618, 1987.

[95]- Williams, R. J. & Zipser, D. *"A learning Algorithm for Continually Running Fully Recurrent Neural Networks"* Neural Computation, Vol.1, pp.270-280, 1989.

[96]- Pearlmutter, B. A. *"Dynamic Recurrent Neural Networks"* Report CMU-CS-88-91,School of Computer Science, Carnegir Mellon University, Pittsburgh, PA, 1988.

[97]- Zipser, D. *"A Subgroubing strategy that Reduces Complexity and speeds Up Learning in recurrent Networks"* Neural Computation, Vol.1, pp.552-558, 1989.

[98]- Williams, R. J. & Peng, J. *"An Efficient Gradient-Based Algorithm for On-Line Training of Recurrent Network Traj*ectories" Neural Computation, Vol.1, pp.270-278, 1989.

[99]- Atiya, A. F. *"Learning on a General Network, in Neural Information Processing Systems"* Anderson, D. Z., American Institute of Physics, New York, 1988.

[100]- Rumelhart, D. E. & Hinton, G. E. & Williams, R. J. "Learning Internal Representations by Error propagation" In Rumelhart, D. E. & McClelland,J. L.,eds., *Prallel Distributed Processing,* Vol.1, Chpter 8. Reprinted in Anderson & Rosenfeld [1988], pp.675 - 695, 1986 -a.

[101]- Rumelhart, D. E. & Hinton, G. E. & Williams, R. J. *"Learning Representations by Error propagation"* Nature, Vol.323, pp.533-536. Reprinted in Anderson & Rosenfeld[1988], pp.696-699, 1986-b.

[102]- Elman, J. L. *"Distributed representations, Simple Recurrent networks ,and Grammatical Structure, Machine Learning"* Vol.7, pp. 195- 225,1991.

[103]- Servan-Schreiber, D; & Cleeremans, A. & McClelland, J. L. *"Graded State Machines: The representation of temporal Contingencies in Simple recurrent networks"* Machine Learning, Vol.7, pp.161-193, 1991.

[104]- Sterzing, V. & Schurmann, B. *"Recurrent Neural Networks for Temporal Learning of Time Series"* Proceedings of the IEEE International Comference on Neural Networks, San Francisco, Vol.2, pp. 843-846, 1993.

[105]- Li, Liang & Haykin, S. *"A Cascade Recurrent Neural Network for Real-Time Nonlinear Adaptive Filtering"* Proceedings of the IEEE International Conference on Neural Networks, San Francisco, pp.857-862,1993.

[106]- Mori, H. & Ogasawara, T. *"A Recurrent Neural Network Approach to Short-Term Load Forecasting in Electronic Power Systems"* Proceedings of the World Congress on Neural Networks, Portland, OR, Vol.I, pp.342-345, 1993.

[107]- Rao, S. S. & Rammamurt, V. *"A Hybrid Technique to Enhance the Performance of Recurrent Neural Networks for Time Series Prediction"* Proceedings of the IEEE International Conference on Neural Networks, San Francisco, 1993.

[108]- Freisleben, B. *"The Composer: A Network for Musical Appications, in Artificial Neural Networks"* Vol.2, Aleksander, I. & Taylor, J., eds., North-Holland, Amesterdam, pp.1663-1666, 1992.

[109]- Fernande, S. & Islam, F. & Utama, P. & Watson, K. *"High Impedance Fault Detection Using recurrent Network, Artificial neural Networks"* Vol.2,

Aleksander, I. & Taylor, J., eds., North-Holland, Amsterdam, pp.1615-1618, 1992.

[110]- Hoshino, T & Kano, M. & Endo, T. *"Optical Control with a Recurrent Network and a priori Knowledge of the System"* Proceedings of the IJCNN-91, Singapore, pp.226-231, 1991.

[111]- Imai, K. *"Simple Recurrent Neural Networks Applied to the Recognition of a Lateral String of Letters"* Private communication between l'auther and Reference[18], unpublished paper, 1991.

[112]- Wang, P. Z. "Truth-Valued Flow Inference Theory and its Application, in Advances in Fuzzy Systems: Application and Theory" Wang, P. Z. & Loe, K. F., eds., World Scientific, Singapore, 1993.

[113]- Hinton, G. E. & Sejnowski, T. J. "Analyzing Cooperative Computation" Proceedings of the Fifth Annual Conference of the Cognitive Science Society, Rochester, NY, pp.448-453, 1989.

[114]- Ackley, D. H. & Hinton, G. E. & Sejnowski. T. J. *"A Learning Algorithm for Boltzmann Machines"* Cognitive Sience,Vol.9,pp.147-69,1985.

[115]- Fausett, L. *"Fundamentals of Neural Networks: Architecturesm Algorithm, Applications"* Prentice Hall, Englewood Cliffs, NJ,1994.

[116]- Gemen, S. & Gemen, D. *"Stochastic Relaxtation, Gibbs Distribhtions and the Bayesian Restoration of Imaes"* IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6, pp.721-741, 1984.

[117]- Szu, H. & Hartly, E. Physics Letters, pp.157-162, 1987.

[118]- Garey, M. & Johnson, D. *"Computers and Intractability: A Guide to the Theory of NP-Completeness"* Freeman, W. H., San Francisco, 1979.

[119]- Light, L. W. & Anderson, P. *"Designing Better Keyboards via Simulated Annealing"* AI Expert, pp.20-7,September, 1993.

[120]- Aarts, E. & Korst , J. *"Simulated Annealing and Boltzmann machines : A Stochastic Approach to Combinatorial Optimization and Neural Computing"* Wiley, New York, 1989.

[121]- Lawler, E. L. & Lenstra, J. K. & Rinnooy Kan, A. H. G.& Shmoys, D. B. *"The Traveling Salesman Problem: A Guided Tour of Combinatorial Optimization"* New York, John Wiley & Sons, 1985.

[122]- Wilson, G. V. & Pawley, G. S. *"On the stability of the Traveling Salesman Problem Alogrithm of Hopfield and Tank"* Biological Cybernetics, Vol.58, pp.63-70, 1988.

[123]- Szu, H. H. *"Fast TSP Algorithm Based on Binary Neuron Output and Analog Neuron Input Using the Zero-Diagonal Interconnect Matrix and necessary and Sufficient Constraints of the Permutation Matrix"* IEEE International Conference on Neural Networks, San Diago, CA, Vol.II, pp.259-265, 1988.

[124]- Hopfield, J. J. & Tank, D. W. *"Neural Computation of Decisions in Optimization Problems"* Biological Cybernetics, Vol.52, pp.141-152, 1985.

[125]- Takefuji, Y. *"Neural network Parallel Computer"* Boston: Kluwer Academic Publishers, 1992.

[126]- Croall, I. F. & Mason, J. P. (eds.) *"Industrial Applications of neural networks"* In project ANNIE Handbook, Springer-Verlag, Berlin, 1991.

[127]- Moallemi, C. *"classifying Cells for Cancer diagnosis Using Neural Networks"* IEEE Expert, December, pp.8-12, 1991.

[128]- Sone, Tadashi *"Using Distributed Neural Networks to Identify Faults in Switching Systems"* Proceedings of the International Workshop on Applications of Neural Networks to telecommunications, Alspector, j. & Goodman, R.& Brown, T. X., Lawrence Erlbaum associates Hillsdale, NJ, 1993.

[129]- Abe, Shigeo & Kayama, M. & Taenaga, H. *"synthesizing neural networks for Pattern Recognition"* Proceedings of the IJCNN-9I, Singapore, pp.1105-10, 1991.

[130]- Caudill, M. *"Neural Networks Primer"* San Francisco, Miller Freeman, 1989.

[131]- Abu-Mostafa, Y. S. *"Vapnik-Chervonenkis Dimension : Information Versus Complexity in Learning"* Neural Computation, Vol.1, pp312-7, 1989.

[132]- Zadeh, L. A. *"Fuzzy Logic, Neural Networks , Soft Computing"* Communications of the ACM,Vol.37, No.3, pp.77-84, 1994.

[133]- Jahne, B. *"Digital Image Processing"* Springer-Verlag, Berlin, Germany, 1997

[134]- Ackley, D. H. "*A Connectionist Machine for Genetic Hillclimbing* " Kluwer Academic Publishers, Boston, 1987.

[135]- Barhin,J. & Gulati, S. & Zak, M. "*Neural learning of constrained Nonlinear Transformation*" Computer, Vol.22(6), pp.67-76, 1989.

[136]- Youssef, H. M. "Comparison of Neural Networks in Nonlinear System Modeling" proceedings of the World Congress on Neural Networks, Portland, OR, pp.IV5-9, 1993.

[137]- Weigend, A. S. & Gershenfeld, N. A. "*Time Series Prediction: Forecasting the Future and Understanding the Past*" Addition-Wesley, Reading, MA, 1994.

[138]- Carpenter,G. A. & Grossberg, S. "*The Art of Adaptive Pattern Recognition by a Self-Organizing Neural Network*" Computer,Vol.21, pp.77-88, 1988.

[139]- Carpenter,G. A. & Grossberg, S.& Rosen, D. B. "*Fuzzy Art: Fast Stable Learning and Categorization of Analog Input Patterns by an Adaptive Resonance System*" Neural Networks, Vol.4, pp.759-71, 1991.

[140]- Ahmad, S. & Tesauro, G. "*Scaling an Generalization in Neural networks*" In D. S. Touretzky, ed., Advances in Neural Information Processing Systems1, San Mateo, CA: Morgan Kaufmann, pp.160-168, 1989.

[141]- Szu, H. & Liu, K. W. & Chao, C. C. & Lin, K. F. & Hsu, K. T. & Medsker, L. Proceedings of the IJCNN-92, Beijing, pp.I-333,I-339, 1992.

[142]- Vemuri, V. "*Artificial Neural Networks: Theoretical Concepts*" Washington, DC: IEEE Computer Society Press, 1988.

[143]- Akiyama, Y. & Yamashita, A. & Kajiura, M. & Aiso, H. "*Combinatorial Optimization with Gaussian Machines*" International Joint Conference on Neural networks, Washington, DC, I:533-540, 1989.

[144]- Ritter, H. J. & Martinetz, T. & Schulten, K. J. "*Neural Computation and Self-Organizing Maps: An Introduction*" Addison-Wesley, Readings, MA, 1992.

[145]- Specht, D. F. "*Probabilistic Neural Networks*" Neural networks,Vol.3, pp.109-18,1990.

[146]- Xu, L. & Qja, E. & Susen, G. S."*Modified Hebbian Learning for Curve and Surface fitting*" Neural Networks, Vol.5(3), pp.441-457, 1992.

[147]- Yager, R. R. *"Modeling and Formulating Fuzzy Knowledge Bases Using Neural Networks"* Neural Networks,Vol. 7, No.8, pp.1273- 83, 1994.

[148]- Alman, W. f. *"Apprentices of Wonder: Inside the Neural Network Revolution"* New York, Bantam Books, 1989.

[149]- Baum, E. B. & Hausler, D. *"What Size Net Gives Valid Generalization?"* Neural Computation,Vol.1, pp.151-60, 1989.

[150]- Block, H. D. *"The Perceptron : A Model for Brain Functioning"* Reviews of Modern Physics, Vol.34, pp.123-35, 1962.

[151]- Blum, E. B."A Proposal for More Powerful Learning Algorithms" Neural Computation,Vol.1, pp.201-207, 1989.

[152]- Rade, L. & Westergren, B. *"Beta Mathematics Handbook"* Studentlitterratur, Chartwell0Bratt Ltd., 1990.

[153]- Caudill. M. & Butler, C. *"Naturally Intelligent Systems"* Cambridge, MIT Press, MA, 1990.

[154]- Dahl,E. D. *"Accelerated learning Using the Generalized Delta rule"* Proceedings of the First IEEE International Conference on Neural networks, San diago, 1987.

[155]- Murat Tekalp, A. *"Digital Video Processing"* Prentice-Hall, Simon & Schuster, NJ, 1995.

[156]- Oja,E. *"Principal Components , Minor Components , and Linear Neural Networks"* Neural Networks, Vol.5(6), pp.927-935,1992.

[157]- Miller, W. T. & Sutton, R. S. & Werbos, P. J. *"Neural Networks for control"* Cambridge , MIT Press, eds., MA, 1990.

[158]- Murry,D. "Tuning Neural Networks with Genetic Algorithms" AI Expert, pp.27-32, June, 1994.

[159]- Almeida, L. B., *"Backpropagation in perceptrons with Feedback "*In R. Eckmiller, & CH. Von der Malsburg, eds., Neural Computers. Berlin, Springer-Verlag, pp.199-208, 1988.

[160]- Amari,S-I *"A Theory of Adaptive Pattern Classifiers "* EEE Transactions on Electronic Computers,Vol.EC, pp.299-307, 1967.

[161]- Chan, S. C. & Hsu, L. S. & Loe, K. F. & The, H. H. *"Neural Logic networks"* Internel Publication of the National University of Singapore, Singapore, pp.1-54.1991.

[162]- Chang, C. F. & Sheu, B. & Thomas, J. *"Multilayered Backpropagation Neural Networks for Financial Analysis"* Proceedings of the INNS Meeting, World Congress on Neural Networks , Portland, OR,Vol.I,pp.445-50,1993.

[163]- Amari, S-I & Fujita, N. & Shinomoto, S. *"Four Types of Learning Curves"* Neural Computation,Vol.4, pp.605-618, 1992.

[164]- Levine, D. S. *"Introduction to Neural and Cognitive Modeling"* Hillsdale, NJ: Lawrence Erlbaum Associates, 1991.

[165]- MacGregor, R. J. *"Neural and Brain Modeling"* San Diago, Acadenic Press, 1987.

[166]- McCoulloch, W. S. *"Embodiments of mind"* Cambridge, MIT Press, MA, 1988.

[167]- Barr,D. S. & Mani, G. *"Using Neural Nets to Manage Investments"* AI Expert,Vol.9, No.2, pp.16-21, 1994.

[168]- Kosko, B. *"Neural Networks and Fuzzy Systems; A dynamical Systems Approach to Machine Intelligence"* Englewood Cliffs, NJ, Prentice-Hall, eds., 1992-a.

[169]- Kosko,B. *"Neural networks for signal Processing"* Englewood Cliffs, NJ, Prentice-Hall, eds., 1992-b.

[170]- Lawrence, J. *"Data preparation for a neural Network"* Neural Network Special report, AI Expert, pp. 15-21, 1993.

[171]- Anderson , J. A. *"A memory Model Using Spatial Correlation Functions"* Kybernetik,Vol.5, pp.113-9, 1968.

[172]- Anderson , J. A. *"Two models for Memory Organization"* Mathematical Biosciences, Vol.8, pp.137-60, 1970.

[173]- Anderson , J. A. & Rosenfeld, E. *"Neurocomputing: Foundations of Research"* Cambridge, MIT press, MA, 1988.

[174]- Anderson , J. A. & Rosenfeld, E. *"Neurocomputing2: Directions for Research"* Cambridge, MIT press, MA, 1990.

[175]-Leondes, Cornelius T. *"Neural Network Systems, Techniques and Applications"* ACADEMIC PESS,New York, 1998

[176]- Angeniol, B. & De La Croix Vaubois, G. & Le Texier, J.-Y. *"Self-Organizing Feature Maps and the Traveling Salesman Problem"* Neural Networks,Vol.1, pp.289-93, 1988.

[177]- Kohonen, T. *"How to Make a Machine transcribe Speech?"* In Applications of neural Networks, H. G. Schuster(ed.), VCH Verlags-gesellschaft, Weinheim, 1992.

[178]- Kohonen, T. *"Self-Organization and Associative Memory"* Second edition,Springer-Verlag, Berlin-Heidelberg, 1987.

[179]- Kohonen, T. *"Neural Phonetic Typewriter"* IEEE Computer,Vol.21, No.3, pp.11-22, 1988.

[180]- Bornholdt, S. & Graudez, D. *"General Asymmetric Neural Networks and Structure Design by Genetic Algorithms"* Neural Networks, Vol.5, pp.327-34, 1992.

[181]- Chiuh,T. D. & Tang, T. T. & Chen, L. G. *"Vector Quantization Using Tree-Structured Self-Organizing Feature Maps"* Proceedings of the International Workshop on Applications of Neural Networks to Telecommunication, J. Alspector & R. Goodman & T. X.Browk(eds.), Lawrence Erlbaum Associstes, Hillsdale, NJ, pp.259-65, 1993.

[182]- Cohen, M. A. & Tesauro, G. *"How Tight are the Vapnik-Chervonenkis bounds?"* Neural Computation,Vol.4, pp.249-69, 1992.

[183]- DARPA *"DARPA neural Network Study"* Final Report, Cambridge, MA:Massachusetts Institute of technology, Lincoln Laboratory, 1988.

[184]- Fukushima, K. *"A Neural Network for Visual Pattern Recognition"* IEEE Computer,Vol. 21, No. 3, pp. 65-75, 1988.

[185]- Fukushima, K. & Wake, N. *"Handwritten Alphabetic Character Recognition by the Neocognitron"* IEEE Transactions on Neural Networks, Vol.2, No.3, pp.355-65, 1991.

[186]- Geman, S. E. & Bienenstock, E. & Doursat, R. *"Neural Networks and the Bias/Variance Dilemma"* Neural Computation,Vol.1, pp.1-58, 1992.

[187]- Burrascano, P. *"Learning Vector Quantization for the probabilistic Neural Network"* IEEE Transactions on Neural Networks, Vol.2, No.4, pp.458-61, 1991.

[188]- Dayhoff,J. E. "*Neural network Architectures*" New York, VanNostrand Reinholt, 1990.

[189]- Gruber, S. L. & Villalabos, L. & Olsson, J. *"Neural Networks for Webb-Process Inspection"* Proceedings of the SPIE Applications of Artificial neural Networks IV, pp.491-503, 1993.

[190]- Harston, C. T. *"Business with Neural networks"* In A. J. Maren & C. T. Harston & R. M. Papeds., Handbook of Neural Computing Applications. San Diago: Academic Press, pp391-400, 1990.

[191]- Drago, G. P. & Ridella, S. *"Cascade Correlation : An Incremental Tool for Function Approximation, in Neural Information Processing Systems"* Vol.2, R. P. Lippman, J. E. Moody & D. S. Touretzky, eds., Morgan Kaufmann, San Mateo, CA, pp.750-756, 1991.

[192]- Fausett, L. *"Fundamentals of Neural Networks"* Prentice-Hall, Englewood Cliffs, NJ, 1994.

[193]- Kasuba, T. *"Simplified Fuzzy ARTMAP"* AI Expert, Vol.8, No.11, pp.18-25, 1993

[194]- Goldberg, D. E. *"Genetic Algorithms"* Addison-Wesley, Reading, MA, 1989.

[195]- Hecht-Nielsen, R. *"Theory of the Backpropagation Neural Network"* International Joint Conference on Neural networks, Washington, DC, Vol. I, pp.593-605, 1989.

[196]- Freedman, D. H. *"Brainmakers"* Simon & Schuster, New York, 1994.

[197]- Grenender, U. *"Abstract Inference"* Wiley, New York, 1981.

[198]- Grossberg, S. *"Studies of Mind And brain"* Boston, Reidel, 1982.

[199]- Hopfield, J. J. & Tank, D. W. *"Computing with Neural circuits"* Science,Vol. 233, pp.625-633, 1986.

[200]- Karr, Chuck *"Applying Genetics to Fuzzy Logic"* AI Expert, pp. 38 - 43, March, 1991.

[201]- Cover, .M. T. & Thomas, J. A. *"Elements of Information Theory"* John Wiley & Sons,USA, 1991.

[202]- Reily, D. L. & Cooper, L. N. & Elbaum, C. *"A Neural Model for Category Leaning"* Biological Cybernetics, Vol.45,pp.34-51,1982.

[203]- Reily, D. L. & Cooper, L. N. & Elbaum, C. *"Learning System Architectures Composed on Multiple Learnin Modules"* Proceedings of the IEEE International Conference on Neural Networks,Vol.2,pp.495-503,1987.

[204]- Scofield, C. L. *"Learning Internal Representations in the Coulomb Energy Network"* IEEE International Conference on Neural Network, San Diego ,CA,Vol.I,pp.271-276,1988.

[205]- Fahlman, S. E. & Labiere, C. *"The Cascace-Correlation Learning Architecture"* Carnegie MillonReport,No. CMU-CS-88-162,1990.

[206]- Littman,E. & Ritter, H. *"Cascade Network Architectures"* Proceedings of the International Joint Conference on Neural Networks, Baltimore, Vol. II, pp. 398 - 404 , 1992.

[207]- Drago, G. P. & Ridella, S. *"An Optimum Weights Initializations for Improving Scaling Relationships in BP learning in Artificial Neural Networks"* Vol.1, Kohonen, T. & Makisara, O. Simula & Kangas,J., eds., North - Holland, Amsterdam, pp.1519 - 1522,1991.

[208]- Gallant, S. I. *"Connectionist Learning Algorithm with Provable Generalization and Scaling Bounds"* Neural Networks,Vol.3, pp . 191-201,1990.

[209]- Frean, M. *"The Upstart Algorithm: A Method for Constructing and Training Feedforward Neural Networks"* Neural Computation,Vol.2, pp. 198 – 209 , 1990.

[210]- Li, Wei & Nasrabadi, M. *"Invariant Object Recognition Based on a Neural Network of Cascade RCE Nets"* Proceedings of the IJCNN - 90, San Diego,Vol.2, pp. 845 - 854, 1990.

[211]- Hasegawa, A. & Shibata, K. & Itoh, K. &Ichioka, Y,Inamura, K. *"Adapting- Size Neural Network for Character Recognition on X-Ray Films"* Proceedings of the International Workshop on Application of Neural Networks to telecommunications, Alspector, J. & Goodman, R. & Brown, T. x. ,eds., Lawrence Elbaum Associates,Hillsdale,NJ,pp.139-146,1993.

[212]-Fukushima, K. & Miyake, S. & Ito, T. *"neocognitron: A Neural Network Model for a Mechanism f Visual Pattern Recognition"* IEEE transactions on Systems, Man, and Cybernetics, Vol. SMC -13, pp. 826-834, 1983.

[213]- Fukushima, K. *"Cognitron: A Self-Organizing Multilayered Neural Network"* Biological Cybernetics ,Vol. 20, pp. 121-136,1975.

[214]- Specht, D. F. *"probabilistic Neural Networks for Classifications, Mapping or Associative Memory"* Proceedings of the IEEE International Conference on Neural Networks, San Diego,Vol.1,pp.525-532,1988

[215]- Mood,A. M. & Graybill, F. A. *"Introduction to the Theory of statistics"* MacMillan, New York.

[216]- Parzen, E. *"On Estimation of a Probability Density Function and Mode"* Annalas of Mathematical Statistis,Vol.33, pp.1065-1076 ,1962.

[217]- Cacoullos, T. *"Estimation of a Multivariate Density"* Annals of Institude of statistical Mathematics,Vol.18, No.2, pp.179-189,1966.

[218]- Specht, D. F. *"A General Regression Neural Network"* IEEE Transactions on Neural Networks, Vol.2, No.6, pp.568-576, 1991.

[219]- Loskiewicz - Buczak, A. & Uhrig, R. E. *"Vibration Data Analysis Using probabilitic Neural Network-Based System"* Proceedings of INNS Meeting, World Congress on Neural Networks, Portland, OR, Lawrence Erlbaum Associates, Hillsdale, NJ, Vol.I, pp.273-278,1993.

[220]- Specht, D. F. *"Generation of Polynomial Discriminant Functions for Pattern Recognition"* IEEE Transactions on Electronic Computers, Vol.EC-16, pp .308-319, 1967.

[221]- Oja, E. *"A Simplified Neuron Model as a Principal Component Analyzer"* Journal of Mathematical Biology, Vol.15, pp.267-273, 1982.

[222]- Oja, E. *"Neural Network,Principal Components,and Subspaces"* International Journal of Neural Systems,Vol.1, pp.61-68, 1989.

[223]- Sanger, T. D. *"An Optimality Principle for Unsupervised Learning, in Advanced in Neural Information Proceeding Systems I"* Tourettzk, D. S., eds., Morgan Kaufmann, San Mateo, CA, pp.11-19, 1989.

[224]- Yuille, A. L. & Kammen, D. M. & Cohen. D. S. *"Quadratic and the Development of orientation Selective Cortical Cells by Hebb Rules"* Bological Cybernetics, Vol.16, pp.183-194, 1989.

[225]- Grossberg, S. *"Neural Expectation: Cerebellar and Retinal Analog Cells Fired by Learnable or Unlearned Pattern Classes"* Kybernetik, Vol.10, pp. 49- 57, 1972.

[226]- Von der Malsburg, C. *"Self-Organization of Orientation Sensitive Cells in the stirate Cortex"* Kybernetik, Vol.14, pp.85-100,1973.

[227]- Kohonnen, T. *"Self-Organization and Associative Memory"* 3rd ed., Berlin, Springer-verlag,1989-a.

[228]- Kohonen, T. *"A Self-learning Musical Grammer, or 'Associative Memory of the Second Kind"* International joint Conference on Neural Networks, Washington,DC, Vol.1, pp, 1-5, 1989-b.

[229]- Kohonen, T. *"Self-Organization and Associative Memory"* First ed., Springer- Verlag, Berlin-Heidelberg,1984.

[230]- DeSieno, D. *"Adding a Consience to Competitive Learning"* Proceedings of the IEEE International Conference on Neural Netorks, San Diego, Vol.1, pp.117-124,1988.

[231]- Kohonen, T. *"Improved Versions of Learning Vector Quantization"* International Joint Conference on Neural Networks,San Diego,CA,Vol.I, pp. 545- 550,1990-a.

[232]- Kohonen, T. *"The Self-Organizing Map"* Proceedings of the IEEE, Vol. 78(9), pp.1464-1480,1990-b.

[233]- Hecht - Nielson, R. *"Counterpropagation Network"* Applied Optics, Vol.26(23), pp.4979-4984,1987-a

[234]- Hecht-Nielson, R. *"Counterpropagation Networks"* IEEE First International Conference on Neural Networks, San Diego,CA,Vol.II, pp.19-32,1987-b.

[235]- Hecht-Nielson, R. *"Kolmogorov's Mapping Neural Network Existence Theorem"* IEEE First International Conference on Neural Networks, San Diego,CA,Vol.III, pp.11-14,1987-c.

[236]- Bogert, B. P. & healy, M. J. R. & Tukey, J. W. *"The Quefrency Analysis of Time Series for Echoes: Cepstrum, Pseudo-autocovariance, Cross-Cepstrum, and Saphe cracking"* Proc.Symposium Time Series Analysis, Rosenblatt, M., ed. John Wiley and Sons, New York, pp.209-243,1963.

[237]- Carpenter, G. A. & Grossberg, S. *" A Massively Prallel Architecture for a Self-Organizing Neural Pattern recognition Machine"* Computer Vision, Graphics, and Image Processing, Vol.37, pp.54-115, 1987-a

[238]- Carpenter, G. A. & Grossberg, S. *"ART2: Self-organization of Stable Category* Recognition Codes for Analog Input Patterns" Applied Optics,Vol.26, pp.4919-4930,1987-b.

[239]- Lippmann, R. P. *"An Introduction to Computing with Neural Nets"* IEEE ASSP Magazine, Vol.4, pp.4-22, 1987.

[240]- Kalkunte, S. S. & Kumar, J. M. & Patnaik, L. M. *"A Neural Networks Approach for High Resolution Fault Diagnosis in Digital Circuits"* Proceedings of he IJCNN-92, Beijing,Vol.I, pp.I-83,I-88, 1992.

[241]- Smith, S. D. G. & Escobedo, R. & Caudekk, T. P. *"An Industrial Strength Neural Network Application"* Proceedings of INNS Meeting, World Congress on Neural Networks, Portland, OR, Lawrence Erlbaum Associate, Hillsdale, NJ, Vol. I, pp. 490 - 494, 1993.

[242]- Teow, L. N. & Lui, H. C. Wang, P. Z. & The, H. H. & Shen, Z. & Goh, T. H. *"Truth Value Flow Inference (TVFI) Neural Network"* private communication with [128]-Patterson, D. W.,1993.

[243]- Hsu, L. S. & The, H. H. & Chan, S. C. & Loe, K. F. *"Fuzzy Decision Making Based on Neural Logic Networks"* Proceedings of the Inter-Faculty Seminar on neuronet Computing, Technical Report DISCs No. Tra-6/89, National University of singapore, Singapore,1989.

[244]- mamdani, E. H. *" Application of Fuzzy Algorithm for Control of Simple Dynamic Plant"* Proceedings of the IEEE, Vol.121, pp. 1585-1589,1974.

[245]- Holland, J. L. *"Adaptation in Neural and Artificial Systems"* University of Michigan press, ANN Arbor, 1975.

[246]- Montana,D. & Davis, L. *"Training Feedforward Networks Using Genetic Algorithms"* Proceedings of the IJCAI-89, Vol.I, pp.762-767,1989.

[247]- Whitley, J. R. & Davis, J. F. *"Qualitative Interpretation of Sensor Patterns"* IEEE Expert, pp.54 - 63, April, 1993.

[248]- Mandischer, M. *"Representation and Evolution of Neural Networks"* IEEE Proceedings of the International Conference on Artificial neural Networks and Genetic Algorithm, Innsbruck, Springer-Verlag, Wien, pp.643-648,1993.

[249]- Mitchell, R. J. & Bishop, J. M. & Low, W. *"Using a genetic Algorithm to Find the Rules of a Neural Network"* IEEE proceedings of the International Conference on Artificial Neural Networks and Genetic Algorithm, Innsbruck, Springer-Verlag, Wien, pp.664-669,1993.

[250]-Teh, A. H. & Tan, A. H. *"Connectionist Expert Systems-A Neural –Logic Model's Approach"* Proceedings of the Inter-Faculty Seminar on neuronet Computing, Technical Report DISCs No. Tra-6/89, National University of Singapore, Singapore, pp.16-32,1989.

[251]-Proakis, J. G. & Rader, C. M. & Ling, F. & Nikias, C. L. *"Advanced digital Signal Processing"* Macmillan publishing company, New York, USA, 1992.




بدمشـــق




السعر : 12 دولار أمريكي أومايعادلها